PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
A61K 39/102, C07K 3/12, 3/18, 3/20, 3/26, 3/28, 15/04, 17/10, C12N 15/31, 15/70, C12P 21/02

A1

(11) International Publication Number: WO 95/03069

(43) International Publication Date: 2 February 1995 (02.02.95)

(21) International Application Number: PCT/US94/08326

(22) International Filing Date: 22 July 1994 (22.07.94)

(30) Priority Data:
08/096,181 23 July 1993 (23.07.93) US

(71) Applicant: NORTH AMERICAN VACCINE, INC. [US/US]; 12103 Indian Creek Court, Beltsville, MA 20705 (US).

(72) Inventors: TAI, Joseph, Y.; 1370 Cinnamon Drive, Fort Washington, PA 19034 (US). PULLEN, Jeffrey, K.; 6928 Garland Lane, Columbia, MA 21045 (US). SOPER, Thomas, S.; Apartment 374, 8216 Gorman Avenue, Laurel, MA 20707 (US). LIANG, Shu-Mei; 6627 River Road, Bethesda, MA 20817 (US).

(74) Agents: ESMOND, Robert, W. et al.; Sterne, Kessler, Goldstein & Fox, Suite 600, 1100 New York Avenue, N.W., Washington, DC 20005-3934 (US).

(81) Designated States: AM, AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, ES, FI, GB, GE, HU, JP, KE, KG, KP, KR, KZ, LK, LT, LU, LV, MD, MG, MN, MW, NL, NO, NZ, PL, PT, RO, RU, SD, SE, SI, SK, TJ, TT, UA, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), ARIPO patent (KE, MW, SD).

Published
With international search report.

(54) Title: METHOD FOR EXPRESSION AND PURIFICATION OF P2 PROTEIN FROM HAEMOPHILUS INFLUENZAE TYPE B

(57) Abstract

The present invention relates, in general, to a method of expressing the outer membrane protein P2 from *Haemophilus influenzae* type b (Hib-P2) and fusion proteins thereof. In particular, the present invention relates to a method of expressing the outer membrane protein P2 from *Haemophilus influenzae* type b or fusion protein thereof in *E. coli* wherein the Hib-P2 protein or fusion protein comprises more than 2 % of the total protein expressed in *E. coli*. The invention also relates to a method of purification and refolding of Hib-P2 protein and fusion protein thereof.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria
AU	Australia
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CS	Czechoslovakia
CZ	Czech Republic
DE	Germany
DK	Denmark
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgystan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakhstan
LI	Liechtenstein
LK	Sri Lanka
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SI	Slovenia
SK	Slovakia
SN	Senegal
TD	Chad
TG	Togo
TJ	Tajikistan
TT	Trinidad and Tobago
UA	Ukraine
US	United States of America
UZ	Uzbekistan
VN	Viet Nam

METHOD FOR EXPRESSION AND PURIFICATION OF P2 PROTEIN FROM HAEMOPHILUS INFLUENZAE TYPE B

Background of the Invention

Field of the Invention

The present invention is in the field of recombinant DNA technology, protein expression and vaccines. The present invention relates, in particular, to a method of expressing the outer membrane protein P2 from *Haemophilus influenzae* type b (Hib-P2). The invention also relates to a method of purification and refolding of the recombinant protein.

Background Information

Haemophilus influenzae type b causes bacterial meningitis and other invasive infections in children under the age of 4 years in the United States. The P2 protein from several *H. influenzae* type b strains has been purified and characterized (Munson *et al., J. Clin. Invest. 72*:677-684 (1983) and Vachon *et al., J. Bacteriol. 162*:918-924 (1985)). The structural gene encoding the P2 protein type 1H has been cloned and the DNA sequence determined (Hansen, E.J., *et al., Infection and Immunity 56*:2709-2716 (October 1988); Hansen, E.J., *et al., Infection and Immunity 57*:1100-1107 (April 1989); and Munson, Jr., R., and Tolan, Jr., R.W., *Infection and Immunity 57*:88-94 (January 1989)).

Although recombinant P2 genes have been expressed in *H. influenzae* Rd (Hansen, E.J., *et al., Infection and Immunity 56*:2709-2716 (October 1988)) and in *E. coli* (Munson, Jr., R., and Tolan, Jr., R.W., *Infection and Immunity 57*:88-94 (January 1989)), the level of expression present in *E. coli* was low, possibly due to the toxicity of the P2 protein in *E. coli* as suggested by Munson (Munson, Jr., R., and Tolan, Jr., R.W., *Infection and Immunity 57*:88-94 (January 1989)) and Hansen (Hansen, E.J., *et al., Infection and*

Immunity 56:2709-2716 (October 1988)). The present invention provides a method of expressing Hib-P2 in *E. coli* wherein the Hib-P2 protein comprises more than 2% of the total protein expressed in *E. coli*.

Summary of the Invention

It is a general object of the invention to provide a method of expressing recombinant outer membrane protein P2 from *Haemophilus influenzae* type b (Hib-P2), or a fusion protein thereof, in *E. coli*.

It is a specific object of the invention to provide a method of expressing the outer membrane protein P2 from *Haemophilus influenzae* type b (Hib-P2), or a fusion protein thereof, in *E. coli* comprising:

(a) transforming *E. coli* by a vector comprising a selectable marker and gene coding for a protein selected from the group consisting of

(i) a mature P2 protein and

(ii) a fusion protein comprising a mature P2 protein fused to amino acids 1 to 22 of the T7 gene ϕ10 capsid protein;

wherein said gene is operably linked to the T7 promoter; and

(b) growing the transformed *E. coli* in LB media containing glucose and a selection agent at about 30°C,

wherein the protein so produced comprises more than 2% of the total protein expressed in the *E. coli*.

It is another specific object of the invention to provide a method of purifying and refolding an outer membrane protein P2 from *Haemophilus influenzae* type b (Hib-P2), or a fusion protein thereof, produced according to the above-described methods.

It is a further specific object of the invention to provide a vaccine comprising the outer membrane protein P2 from *Haemophilus influenzae* type b (Hib-P2), or a fusion protein thereof, produced according to the above methods, in an amount effective to elicit protective antibodies in an animal to

Haemophilus influenzae type b; together with a pharmaceutically acceptable diluent, carrier, or excipient.

It is another specific object of the invention to provide the above-described vaccine, wherein said outer membrane protein P2 or fusion protein thereof is conjugated to a *Haemophilus* capsular polysaccharide.

It is a further specific object of the invention to provide a method of preventing bacterial meningitis in an animal comprising administering to the animal the Hib-P2 protein or fusion protein-vaccine produced according to the above-described methods.

It is another specific object of the invention to provide a method of preparing a polysaccharide conjugate comprising: obtaining the above-described outer membrane protein P2 or fusion protein; obtaining a polysaccharide from a *Haemophilus* organism; and conjugating the protein to the polysaccharide.

It is another specific object of the invention to provide a method of purifying the above-described outer membrane protein P2 or fusion protein comprising: lysing the transformed *E. coli* to release the P2 protein or fusion protein as part of insoluble inclusion bodies; washing the inclusion bodies with a buffer to remove contaminating *E. coli* cellular proteins; resuspending and dissolving the inclusion bodies in an aqueous solution of a denaturant; diluting the resultant solution in a detergent; and purifying the solubilized P2 protein or fusion protein by gel filtration.

It is another specific object of the invention to provide a method of refolding the above-described outer membrane protein P2 or fusion protein comprising: lysing the transformed *E. coli* to release the P2 protein or fusion protein as part of insoluble inclusion bodies; washing the inclusion bodies with a buffer to remove contaminating *E. coli* cellular proteins; resuspending and dissolving the inclusion bodies in an aqueous solution of a denaturant; diluting the resultant solution in a detergent; and purifying the solubilized P2 protein or fusion protein by gel filtration; and storing the gel filtration product

at about 4°C in an aqueous solution containing high concentrations of NaCl and calcium ions until the outer membrane protein P2 refolds.

Further objects and advantages of the present invention will be clear from the description that follows.

Brief Description of the Drawings

FIGURE 1. *Electrophoretic gel showing the kinetics of induction of plasmid pNV-3.* (Coomassie blue stained linear 8-16% gradient SDS-PAGE (Novex)). Lane 1 shows molecular weight markers: phosphorylase b (97.4 kDa), bovine serum albumin (66 kDa), ovalbumin (45 kDa), carbonic anhydrase (31 kDa), soybean trypsin inhibitor (21.5 kDa) and lysozyme (14.4 kDa). Lanes 2 and 14 show 4 μg samples of purified rHib porin (recombinant). Lanes 3-13 show samples of *E. coli* extracts obtained from cells removed at 0, 15, 30, 45, 60, 120, 180, 240, 300, 360 and 420 minutes after addition of IPTG to the culture. At each time point, 5 ml of the culture was removed and immediately chilled to 4°C. The cells were then collected by centrifugation and stored at -75°C. A whole cell extract was made by adding 150 μl of Tris-HCl, pH = 8.0, 5 M urea, 1% SDS, 30 mM NaCl, 2.5% β-mercaptoethanol and 0.05% bromphenol blue. After boiling the mixture for 5 minutes, the samples were then diluted 1:10 with load buffer and then 10 μl of the diluted sample loaded per lane.

FIGURE 2. *Electrophoretic gel showing the kinetics of induction of plasmid pNV-6.* (Coomassie blue stained linear 8-16% gradient SDS-PAGE (Novex)). Lane 1 shows molecular weight markers: phosphorylase b (97.4 kDa), bovine serum albumin (66 kDa), ovalbumin (45 kDa), carbonic anhydrase (31 kDa), soybean trypsin inhibitor (21.5 kDa) and lysozyme (14.4 kDa). Lanes 2 and 14 show 4 μg samples of purified rHib porin. Lanes 3-13 show samples of *E. coli* extracts obtained from cells removed at 0, 15, 30, 45, 60, 120, 180, 240, 300, 360 and 420 minutes after addition of IPTG to the culture. At each time point, 5 ml of the culture was removed and immediately

chilled to 4°C. The cells were then collected by centrifugation and stored at -75°C. A whole cell extract was made as described in Figure 1.

FIGURES 3 and 3A. *A graph showing the gel filtration of rHib porin.* Inclusion bodies were extracted with 6 M guanidine HCl and detergent was added as described in Example 6. The mixture was centrifuged to remove any residual material and applied to a 180 x 2.5 cm S-300 column equilibrated in 100 mM Tris-HCl, 10 mM EDTA, 1 M NaCl and 0.05% 3,14-Zwittergent, at pH 8.0. A second batch was then applied in the same buffer with 20 mM $CaCl_2$. The optical density at 280 nm was measured for each fraction. The arrows indicate the elution position of molecular weight markers (Sigma); 1 = blue dextran (2,000 kDa), 2 = alcohol dehydrogenase (150 kDa); 4 = bovine serum albumin; and 6 = cytochrome C (12.4 kDa). The insert shows a semilog plot of apparent molecular weight versus the elution position. Number 3 is the position of the major peak of the calcium ion treated porin, while number 5 is the position of the major peak of the untreated porin.

FIGURES 4A-4C. *The DNA sequence of the SalI-SalI fragment of pNV-1.* Restriction sites are underlined. The synthetic oligonucleotides used to sequence the DNA are shown doubly underlined. The arrows indicate the direction of the sequencing reaction. Those with left-arrows are complementary to the shown sequence. The rest of the plasmid is identical to pUC18. The *lac* promotor is adjacent to the lower *Sal*I site.

FIGURES 5A-5C. *The DNA sequence of the BamHI-XhoI fragment of pNV-2.* The portion of the pET-17b vector that encodes the fusion sequence is shown in bold. Restriction sites are underlined. The rest of the plasmid is identical to pET-17b.

FIGURES 6A-6C. *The DNA sequence of the NdeI-XhoI fragment of pNV-3.* Restriction sites are underlined. The rest of the plasmid is identical to pET-17b.

FIGURES 7A-7C. *The DNA sequence of the NdeI-BamHI fragment of pNV-6.* Restriction sites are underlined. The rest of the plasmid is identical to pET-11a.

FIGURES 8A and 8B. *Electrophoretic gel (Panel A) and Western blot (Panel B) showing the immunogenicity of native antiP2 from Haemophilus influenzae with recombinant P2.* Panel A: (Coomassie blue stained, linear 8-16 % gradient SDS-PAGE (Novex)). Lane 1 shows molecular weight markers: phosphorylase b (97.4 kDa), bovine serum albumin (66 kDa), ovalbumin (45 kDa), carbonic anhydrase (31 kDa), soybean trypsin inhibitor (21.5 kDa) and lysozyme (14.4 kDa). Lane 3 shows 1 μg of purified recombinant *H. influenzae* type b porin. Lanes 2, 4, 9, 12 and 15 are blank. Lanes 5 and 6 show *E. coli* strain BL21 before and after 3 hours of induction with IPTG. Lanes 7 and 8 show BL21 [pNV-3] before and after induction. Lanes 10 and 11 show BL21 (DE3) [pNV-3], before and after induction. Lanes 13 and 14 show BL21 (DE3) [pNV-6] before and after induction. The samples loaded were prepared as described for Figure 1 herein. Panel B: (Western blot from a gel loaded in an identical fashion to that shown in Panel A). After transfer of the proteins to the nitrocellulose membrane (Novex), the membrane was blocked with powdered milk. Then, a polyclonal antibody generated by immunization of rabbits with a conjugate vaccine composed of purified P2 from Hib strain A2 which is equivalent to strain Eagan and polysaccharide isolated from the same organism were added. Goat antirabbit IgG coupled to alkaline phosphatase was also added thereafter. Visualization of the porin bands was achieved by using a nitro blue tetrazolium stain (Sigma) that reacted with the released phosphate from 5-bromo-4-chloro-3-indolyphosphate, p-toluidine salt (Sigma) (Blake *et al.*, *Analyt. Biochem. 136*:175-179 (1984)).

Detailed Description of the Invention

The present invention relates to a method of expressing the outer membrane protein P2 from *Haemophilus influenzae* type b or a fusion protein thereof.

In one embodiment, the present invention relates to a method of expressing the outer membrane protein P2 from *Haemophilus influenzae* type b or fusion protein in *E. coli* comprising:

(a) transforming *E. coli* by a vector comprising a selectable marker and a gene coding for a protein selected from the group consisting of

(i) a mature P2 protein and

(ii) a fusion protein comprising a mature P2 protein fused to amino acids 1 to 22 of the T7 gene ϕ10 capsid protein;

wherein said gene is operably linked to the T7 promoter; and

(b) growing the transformed *E. coli* in LB media containing glucose and a selection agent to which *E. coli* is sensitive (preferably, carbenicillin) at about 30°C; whereby the Hib-P2 or fusion protein thereof is expressed,

wherein the Hib-P2 protein or fusion protein thereof so expressed comprises more than about 2% of the total protein expressed in the *E. coli*. In a preferred embodiment, the Hib-P2 protein or fusion protein so expressed comprises more than about 5% of the total protein expressed in *E. coli*. In another preferred embodiment, the Hib-P2 protein or fusion protein so expressed comprises more than about 10% of the total protein expressed in *E. coli*. In yet another preferred embodiment, the Hib-P2 protein or fusion protein so expressed comprises more than about 40% of the total protein expressed in *E. coli*.

In another preferred embodiment, the vector comprises a Hib-P2 gene operably linked to the T7 promoter of expression plasmids pET-17b, pET-11a, pET-24a-d(+) or pET-9a, all of which are commercially available from Novagen (565 Science Drive, Madison, WI 53711). Plasmids pET-17b, pET-9a and pET-24a-d(+) comprise, in sequence, a T7 promoter, a ribosome binding site, restriction sites to allow insertion of the structural gene and a T7 terminator sequence. In addition, pET-11a has a lac operator fused to the T7 promotor and a copy of the *lac*I gene. The plasmid constructions employed

in the present invention are different than those used in Munson, Jr., R., and Tolan, Jr., R.W., *Infection and Immunity* 57:88-94 (January 1989), and allow for an unexpectedly high production of the P2 proteins and fusion proteins.

The transformed *E. coli* are grown in a medium containing a selection agent, e.g. any β-lactam to which *E. coli* is sensitive such as carbenicillin. The pET expression vectors provide selectable markers which confer antibiotic resistance to the transformed organism.

According to the present invention, an extraneous 3′ portion down stream from the P2 gene containing P2 termination sequences is eliminated. The fragment thus constructed ends about 40 bp after the translational stop codon.

Any *E. coli* strain encoding T7 polymerase may be used in the practice of the invention. In a preferred embodiment, *E. coli* strain BL21 (DE3) Δ*omp*A is employed. The above mentioned plasmids may be transformed into this strain or the wild-type strain BL21(DE3). The strain BL21 (DE3) Δ*omp*A is a lysogen of bacteriophage λ DE3, which contains the T7 RNA polymerase gene under the control of the inducible *lac*UV5 promoter. *E. coli* strain BL21 (DE3) Δ*omp*A is preferred as no OmpA protein is produced by this strain which might contaminate the purified porin protein and create undesirable immunogenic side effects. The transformed *E. coli* of the present invention may be grown and induced in LB broth containing glucose and carbenicillin at about 30°C and at a low aeration rate (about 150 rpm). Under these conditions, a high level of P2 expression was obtained.

Long term, high level expression of P2 can be toxic in *E. coli*. The highest expression level of Hib-P2 which has been reported is less than 2% of the total proteins expressed (Munson, Jr., R., and Tolan, Jr., R.W., *Infection and Immunity* 57(1):88-94 (January 1989)). Surprisingly, the present invention allows *E. coli* to express the Hib-P2 protein and fusion protein thereof to a level of about 35-50%, as measured by densitometry on an electrophoresis gel after staining with Coomassie blue.

In another preferred embodiment, the present invention relates to a vaccine comprising the outer membrane protein P2 from *Haemophilus influenzae* type b (Hib-P2) or fusion protein thereof, produced according to the above-described methods, together with a pharmaceutically acceptable diluent, carrier or excipient, wherein the vaccine may be administered in an amount effective to elicit protective antibodies in an animal to *Haemophilus influenzae* type b. In a preferred embodiment, the animal is selected from the group consisting of humans, cattle, pigs, sheep and chickens. In another preferred embodiment, the animal is a human.

In another preferred embodiment, the present invention relates to the above-described vaccine, wherein said outer membrane protein P2 or fusion protein thereof is conjugated to a *Haemophilus* capsular polysaccharide (CP). *Haemophilus* CPs may be prepared or synthesized as described in Schneerson *et al., J. Exp. Med. 152:*361-376 (1980); Marburg *et al., J. Am. Chem. Soc.* 108:5282 (1986); Jennings *et al., J. Immunol. 127:*1011-1018 (1981); and Beuvery *et al., Infect. Immunol. 40:*39-45 (1983); the contents of each of which are fully incorporated by reference herein.

In a further preferred embodiment, the present invention relates to a method of preparing a polysaccharide conjugate comprising: obtaining the above-described outer membrane protein P2 or fusion protein; obtaining a CP or fragment from a *Haemophilus* organism; and conjugating the outer membrane protein P2 or fusion protein to the CP or CP fragment.

The conjugates of the invention may be formed by reacting the reducing end groups of the CP fragment to primary amino groups of the porin by reductive amination. The reducing groups may be formed by selective hydrolysis or specific oxidative cleavage, or a combination of both. Preferably, the CP is conjugated to the porin protein by the method of Jennings *et al.*, U.S. Patent No. 4,356,170, the contents of which are fully incorporated by reference herein, which involves controlled oxidation of the CP with periodate followed by reductive amination with the porin protein.

The vaccine of the invention comprises the Hib-P2 protein, fusion protein or conjugate vaccine in an amount effective depending on the route of administration. Although subcutaneous or intramuscular routes of administration are preferred, the Hib-P2, fusion protein or vaccine of the present invention can also be administered by an intraperitoneal or intravenous route. One skilled in the art will appreciate that the amounts to be administered for any particular treatment protocol can be readily determined without undue experimentation. Suitable amounts might be expected to fall within the range of 2 to 100 micrograms of the protein per kg body weight.

The vaccine of the present invention may be employed in such forms as capsules, liquid solutions, suspensions or elixirs for oral administration, or sterile liquid forms such as solutions or suspensions. Any inert carrier is preferably used, such as saline, phosphate-buffered saline, or any such carrier in which the Hib-P2 protein, fusion protein or conjugate vaccine have suitable solubility properties. The vaccines may be in the form of single dose preparations or in multi-dose flasks which can be used for mass vaccination programs. Reference is made to Remington's *Pharmaceutical Sciences*, Mack Publishing Co., Easton, PA, Osol (ed.) (1980), and *New Trends and Developments in Vaccines*, Voller *et al.* (eds.), University Park Press, Baltimore, MD (1978), for methods of preparing and using vaccines.

The Hib-P2 protein or conjugate vaccines of the present invention may further comprise adjuvants which enhance production of P2 antibodies. Such adjuvants include, but are not limited to, various oil formulations such as Freund's complete adjuvant (CFA), the dipeptide known as MDP, saponin, aluminum hydroxide or lymphatic cytokine.

Freund's adjuvant is an emulsion of mineral oil and water which is mixed with the immunogenic substance. Although Freund's adjuvant is powerful, it is usually not administered to humans. Instead, the adjuvant alum (aluminum hydroxide) may be used for administration to a human. Hib-P2 protein, fusion protein or a conjugate vaccine thereof may be absorbed onto the aluminum hydroxide from which it is slowly released after injection.

Hib-P2 protein, fusion protein or conjugate vaccine may also be encapsulated within liposomes according to Fullerton, U.S. Patent No. 4,235,877.

In another preferred embodiment, the present invention relates to a method of preventing bacterial meningitis in an animal comprising administering to the animal the Hib-P2 protein or conjugate vaccine produced according to methods described in an amount effective to prevent bacterial meningitis.

In a further embodiment, the invention relates to a method of purifying the above-described outer membrane protein P2 or fusion protein, comprising: lysing the transformed *E. coli* to release the P2 protein or fusion protein as part of insoluble inclusion bodies; washing the inclusion bodies with a buffer to remove contaminating *E. coli* cellular proteins; resuspending and dissolving the inclusion bodies in an aqueous solution of a denaturant; diluting the resultant solution in a detergent; and purifying the solubilized P2 protein or fusion protein by gel filtration in the absence of denaturant.

The lysing step may be carried out according to any method known to those of ordinary skill in the art, e.g. by sonication, enzyme digestion, osmotic shock or by passing through a mull press.

The inclusion bodies may be washed with any buffer which is capable of solubilizing the *E. coli* cellular proteins without solubilizing the inclusion bodies comprising the P2 protein or fusion protein. Such buffers include but are not limited to TEN buffer (50 mM Tris HCl, 1 mM EDTA, 100 mM NaCl, pH 8.0). Other buffers can be used such as Bicine, Tricine and HEPES.

Denaturants which may be used in the practice of the invention include 2 to 8 M urea or about 2 to 6 M guanidine HCl, more preferably, 4 to 8 M urea or about 4 to 6 M guanidine HCl, and most preferably, about 8 M urea or about 6 M guanidine HCl.

Examples of detergents which can be used to dilute the solubilized P2 protein or fusion protein include, but are not limited to, ionic detergents such as SDS and Cetavlon (Aldrich); non-ionic detergents such as Tween, Triton

X-100 and octyl glucoside; and zwitterionic detergents such as 3,14-Zwittergent and Chaps.

The solubilized P2 protein or fusion protein may be purified by gel filtration to separate the high and low molecular weight materials. Types of filtration matrices include but are not limited to Sephacryl-300, Sepharose CL-6B and Bio-Gel A-1.5m. The column is eluted with the buffer used to dilute the solubilized protein. The fractions containing the P2 protein or fusion thereof may then be identified by gel electrophoresis, the fractions pooled, dialyzed and concentrated.

Finally, substantially pure (>95%) P2 protein and fusion protein may be obtained by passing the concentrated fractions through a Fast Flow Q Sepharose High Performance (Pharmacia) column.

In another embodiment, the present invention relates to expression of Hib-P2 in a yeast Pichia expression system (Sreekrishna *et al., J. Basic Microbiol. 28:*265-278 (1988)) and an archaebacteria expression system (Blaseio and Pfeifer, *Proc. Natl. Acad. Sci. U.S.A. 87:*6772-6776 (1990); Cline *et al., J. Bacteriol. 171:*4987-4991 (1989)). The cloning of the P2 protein gene or fusion gene into an expression vector may be carried out in accordance with conventional techniques, including blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide appropriate termini, filling in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining and ligation with appropriate ligases. Reference is made to Sambrook *et al., Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press (1989), for general methods of cloning.

The Hib-P2 and fusion protein expressed according to the present invention must be properly refolded in order to achieve a structure which is immunologically characteristic of the native protein. In yet another embodiment, the present invention relates to a method of refolding the above-described outer membrane protein P2 or fusion protein, comprising: lysing the transformed *E. coli* to release the outer membrane protein P2 or fusion protein

as part of insoluble inclusion bodies; washing the inclusion bodies with a buffer to remove contaminating *E. coli* cellular proteins; resuspending and dissolving the inclusion bodies in an aqueous solution of a denaturant; resuspending and dissolving the outer membrane protein P2 in high salt (preferably, 2 to 8 M urea or about 2 to 6 M guanidine HCl, more preferably, 4 to 8 M urea or about 4 to 6 M guanidine HCl, and most preferably, about 8 M urea or about 6 M guanidine HCl); diluting the resultant solution in a detergent (preferably, zwittergent, SDS or Tween-20); purifying the outer membrane protein P2 by gel filtration; and storing the gel filtration product at about 1°C to 15°C (preferably, about 4°C) until the outer membrane protein P2 refolds (preferably, one to 10 weeks; most preferably, about three weeks).

The gel filtration step separates high and low molecular weight material and also allows the separation of trimeric and monomeric porin.

After the gel filtration step, high levels of salt (1 to 4M NaCl) are required initially to keep the porin in solution. Calcium ions (preferably, 1mM to 1M $CaCl_2$; most preferably, about 20mM $CaCl_2$), but not magnesium or manganese ions, are required for efficient aggregation of the rHib porin. At this stage, while the rHib porin is trimeric, the conformation is not "native" because when the salt is removed, the porin precipitates from solution. This does not occur with wild-type Hib porin. However, as the porin is stored at 4°C, a slow conformational change occurs which allows the salt to be removed without precipitation of the porin.

The protein at this stage is about 80 to 90 percent pure as judged by Coomassie blue stained SDS-PAGE. This material is then applied to an ion exchange column and eluted with a salt gradient. The resulting material is ~95% pure.

In another preferred embodiment, the present invention relates to a substantially pure refolded outer membrane protein P2 from *Haemophilus influenzae* type b (Hib-P2) produced according to the above-described methods. A substantially pure protein is a protein that is generally lacking in other cellular *Haemophilus influenzae* components as evidenced by, for

example, electrophoresis. Such substantially pure proteins have a purity of >95% as measured by densitometry on an electrophoretic gel.

The present invention is described in further detail in the following non-limiting examples.

Example 1

Cloning of the Outer Membrane Protein P2 from Haemophilus Influenzae Type B

Total genomic DNA was isolated from 0.5 g of *Haemophilus influenzae* type b strain Eagan using methods previously described (Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press (1989)). This DNA was then used as a template for two P2 specific oligonucleotides in a polymerase chain reaction (PCR) using standard PCR conditions (U.S. Patent No. 4,683,195; U.S. Patent No. 4,683,202; Saiki *et al., Science 230*:1350-1354 (1985); Innis *et al., PCR Protocols: A Guide to Methods and Applications,* Academic Press, Inc., San Diego, CA (1990), the contents of which are fully incorporated by reference herein).

The 5′ P2 specific oligonucleotide was designed to be 40 bp 5′ of the ATG (start codon) and had the sequence (SEQ ID NO:1):

5′ TTC-TGG-CGA-GTC-GAC-AAT-TCT-ATT-GG 3′.

The 3′ P2 specific oligonucleotide was designed to be 300 bp 3′ of the stop codon and had the sequence (SEQ ID NO:2):

5′ AAC-CTT-TAT-CGT-CGA-CGA-GCA-ATT-GG 3′.

Both of the P2 specific oligonucleotides contained *Sal*I restriction enzyme sites to facilitate cloning of the amplified product.

Subsequent to the PCR amplification reaction, the amplified DNA was isolated by electrophoresis on a 0.8% agarose gel. The gel demonstrated a single 1.4 kb band. This DNA was purified from the gel and digested with three restriction enzymes (*Eco*RI, *Dra*I and *Pvu*II) that yielded bands of predictable sizes. The 1.4 kb fragment was then digested with *Sal*I and ligated to *Sal*I digested pUC18 (Yanisch-Perron *et al., Gene 33*:103-119 (1985)) using T4 DNA ligase.

The ligation mixture was used to transform competent *E. coli* strain DH5α (Sambrook *et al., Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989)). Resulting colonies were isolated and then analyzed by preparing mini-prep DNAs. The DNAs were analyzed by digesting with *Sal*I which yielded a vector band of 2.7 kb and a fragment band of 1.4 kb.

The ligation that generated plasmid pNV-1 was nondirectional. This means that the DNA insert should be present in both orientations. To test the orientation of the insert, the plasmid was digested with both *Mlu*I and *Nar*I. The size of the resulting fragments indicates whether the insert is oriented in the same direction as the *lac* promotor, or in the opposite direction. Several isolates of the plasmid were tested and all were found to be in the opposite direction to the *lac* promotor. Evidently, the inserts that were in the same direction as the promotor were selected against during growth. This suggests that expression of the rHib P2 is toxic in *E. coli*. Similar conclusions were reached earlier by Munson's group (Munson and Tolan, *Infect. Immunity 57*:88-94 (1989)) and by Hansen's group (Hansen *et al., Infect. Immunity 56*:2709-2716 (1989)).

Clones containing the 1.4 kb fragment were chosen for DNA sequence analysis. One clone designated pNV-1 was sequenced in both directions using the Sanger method (Sanger *et al., Proc. Natl. Acad. Sci. USA 74*:5463 (1977)). Plasmid pNV-1 was found to be identical to the published sequence for Hib strain Minn A (Munson, Jr., R., and Tolan, Jr., R.W., *Infection and Immunity 57*:88-94 (January 1989)).

Molecular biological techniques used herein may be found in Sambrook *et al., Molecular Cloning: A Laboratory Manual,* 2nd ed., Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press (1989) and Ausubel *et al., Current Protocols in Molecular Biology,* Vols. 1 and 2, Wiley-Liss, New York, NY (1992), the contents of which are fully incorporated by reference herein.

Example 2

Construction of Expression Vectors containing the Outer Membrane Protein P2 gene

The expression vector, pET-17b (Novagen pET System Manual), was used for the expression of P2. This plasmid utilizes the phage T7 ϕ10 gene promotor. This promotor is not recognized by *E. coli* DNA dependent RNA polymerase and thus will not produce substantial levels of the porin unless T7 RNA polymerase is present. Strain BL21 (DE3) contains a lysogenic λ phage that encodes the required polymerase under control of the *lac*UV5 promotor. Two types of recombinant P2 proteins were made using the pET-17b expression vector. One type was the mature P2 containing a methionine at the N-terminus. The second type was a fusion protein (designated fusion-P2) containing the mature P2 with 22 amino acids of gene 10 of phage T7 at the N-terminus that were derived from the pET-17b vector.

To clone the P2 into pET-17b, the original P2 gene (in pNV-1) was modified using PCR. To construct the mature-P2, an oligonucleotide was constructed that allowed the mature porin to be cloned into the *Nde*I site of pET 17b, thus producing the mature-P2. The oligonucleotide designed for this had the sequence (SEQ ID NO:3):

5′ GCT-TCA-GCA-GCA-CAT-ATG-GCT-GTT-GTT-
TAT-AAC-AAC-GAA-GGG-AC 3′.

To construct the fusion-P2, an oligonucleotide was constructed that allowed the mature porin to be cloned into the *Bam*HI site of pET 17b, thus yielding a fusion P2 to gene 10 which is a major capsid protein of T7. The sequence (SEQ ID NO:4) of this oligonucleotide was:

5′ GCA-CT-TCA-GCA-GCG-GAT-CCA-GCT-GTT-GTT-TAT-AAC-AAC-GAA-GGG 3′.

The extraneous 3′ sequences were eliminated by introducing a *xho*I site about 40 bp from the translational stop codon. This oligonucleotide was designed to contain an *Xho*I site to allow it to be cloned into the *Xho*I site of pET-17b. The sequence (SEQ ID NO:5) of this oligonucleotide was:

5′ GC-AAA-AAA-AGC-GAA-TCT-CTC-GAG-TCG-CCT-TGC-TTT 3′.

PCR was used to generate a 1.1 kb fragment from the full length P2 (pNV-1) with the 5′ oligonucleotide containing the *Nde*I site and the 3′ oligonucleotide containing the *Xho*I site. This fragment was digested with *Nde*I and *Xho*I, purified and ligated into *Nde*I-*Xho*I digested pET-17b. This resulted in the mature-P2 construct (pNV-3 or N-X).

Likewise, a 1.1 kb fragment was generated from the full length pNV-1 with the 5′ oligonucleotide containing the *Bam*HI site and the 3′ oligonucleotide containing the *Xho*I site using PCR. This fragment was digested with *Bam*HI and *Xho*I, purified and ligated into the *Bam*HI-*Xho*I digested pET-17b. This yielded the fusion-P2 construct (pNV-2 or B-X). Both of the constructs (pNV-3 and pNV-2) were transformed into *E. coli* DH5α strain which lacks T7 polymerase. Plasmid DNA was isolated from numerous DH5α transformants. Both the mature-P2 and fusion-P2 constructs were sequenced at their 5′ and 3′ ends to ensure that the cloning junctions were correct.

Figure 1 shows the kinetics of induction by IPTG of *E. coli* strain BL21 (DE3) [pNV-3]. Note that even before addition of the gratuitous

inducer, there are significant levels of the porin present. This is because the *lac*UV5 promotor is not fully repressed. The level of porin rapidly increases and reaches a maximum after about three hours.

Porin expression in strain BL21 (DE3) is still toxic. This is due to the significant uninduced levels of the porin observed in Figure 1. Care must be taken in handling this strain (keep frozen when not in use; induce at 30°C) because deletions or other mutations will be selected that do not produce porin.

Example 3

Construction of pNV-6

Plasmid pET-11a (Novagen pET System Manual) has the same expression signals as pET-17b. However, this plasmid also contains the *lac* operator adjacent to the T7 gene ϕ10 promotor. This places the T7 promotor under regulation of the *lac* repressor. Plasmid pET-11a also contains an extra copy of the *lac*I gene that encodes the *lac* repressor. This construction should result in substantially lower uninduced levels of porin.

Plasmid pET-11a contains fewer usable restriction sites than pET17b. There is a *Nde*I site in the same location as in pET17b, thus allowing reuse of oligonucleotide SEQ IN NO:3 at the 5' end of the P2 gene. However, there is no *Xho*I site available. Instead, a *Bam*HI site is incorporated using the oligonucleotide (SEQ ID NO:6):

5' AAA-AAA-AGC-GAA-TCT-TTG-GAT-CCG-CCT-TGC-TTT-TAA-TAA-TG 3'.

PCR was used to generate a new 1.1 kb fragment from full length P2 (pNV-1) with the oligonucleotides 3 and 6. This fragment was digested with *Nde*I and *Bam*HI, purified and ligated into pET11a previously cut with *Nde*I-*Bam*HI. This resulted in a second mature-P2 construct (pNV-6). Both the 5'

and the 3' ends of this construction were sequenced to ensure the cloning junctions were correct.

Figure 2 shows the kinetics of induction of BL21 (DE3) [pNV-6]. Notice that the uninduced levels of the porin are much lower than observed with plasmid pNV-3. The time required to reach the maximum level of induction is slightly longer than observed with pNV-3 but after three hours, the levels of porin are comparable with pNV-3. The lower uninduced levels of porin observed in pNV-6 means that this plasmid should show lower levels of toxicity than plasmid pNV-3 and thus should be more stable.

Example 4

Construction of Expression Strain BL21 (DE3) ΔompA

Escherichia coli strains DME558 (from the collection of S. Benson; Silhavy, T.J., *et al.*, "Experiments with Gene Fusions," Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. (1984)), BRE51 (Bremer, E., *et al.*, *FEMS Microbiol. Lett. 33*:173-178 (1986)) and BL21(DE3) were grown on LB agar plates at 37°C.

P1 Transduction: A Pl_{vir} lysate of *E. coli* strain DME558 was used to transduce a tetracycline resistance marker to strain BRE51 (Bremer, E., *et al.*, *FEMS Microbiol. Lett. 33*:173-178 (1986)) in which the entire *omp*A gene had been deleted (Silhavy, T.J., *et al.*, *Experiments with Gene Fusions*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1984)). Strain DME558, containing the tetracycline resistance marker in close proximity of the *omp*A gene, was grown in LB medium until it reached a density of approximately 0.6 $OD_{600\ nm}$. One tenth of a milliliter of 0.5 M $CaCl_2$ was added to the 10 ml culture and 0.1 ml of a solution containing 1 x 10^9 PFU of $P1_{vir}$. The culture was incubated for 3 hours at 37°C. After this time, the bacterial cell density was visibly reduced. One-half of a milliliter of chloroform was added and the phage culture stored at 4°C. Because typically 1-2% of the *E. coli* chromosome can be packaged in each phage, the number

of phage generated covers the entire bacterial host chromosome, including the tetracycline resistance marker close to the *ompA* gene.

Next, strain BRE51, which lacks the *ompA* gene, was grown in LB medium overnight at 37°C. The overnight culture was diluted 1:50 into fresh LB and grown for 2 hr. The cells were removed by centrifugation and resuspended in MC salts. One-tenth of a milliliter of the bacterial cells were mixed with 0.05 of the phage lysate described above and incubated for 20 minutes at room temperature. Thereafter, an equal volume of 1 M sodium citrate was added and the bacterial cells were plated out onto LB plates containing 12.5 μg/ml of tetracycline. The plates were incubated overnight at 37°C. Tetracycline resistant (12 μg/ml) transductants were screened for lack of OmpA protein expression by SDS-PAGE and Western blot analysis, as described below. The bacteria resistant to the antibiotic have the tetracycline resistance gene integrated into the chromosome very near where the *ompA* gene had been deleted from this strain. One particular strain was designated BRE-T^R.

A second round of phage production was then carried out with the strain BRE-T^R using the same method as described above. Representatives of this phage population contain both the tetracycline resistance gene and the *ompA* deletion. These phage were then collected and stored. These phage were then used to infect *E. coli* BL21(DE3). After infection, the bacteria contained the tetracycline resistance marker. In addition, there was a high probability that the *ompA* deletion was selected on the LB plates containing tetracycline.

Colonies of bacteria which grew on the plates were grown up separately in LB medium and tested for the presence of the OmpA protein. Of those colonies selected for examination, all lacked the OmpA protein as judged by antibody reactivity on SDS-PAGE western blots.

The SDS-PAGE was a variation of Laemmli's method (Laemmli, U.K., *Nature 227*:680-685 (1970)) as described previously (Blake and Gotschlich, *J. Exp. Med. 159*:452-462 (1984)). Electrophoretic transfer to Immobilon P

(Millipore Corp. Bedford, MA) was performed according to the methods of Towbin *et al.* (Towbin, H., *et al.*, *Proc. Natl. Acad. Sci. USA 76*:4350-4354 (1979)) with the exception that the paper was first soaked in methanol. The Western blots were probed with phosphatase conjugated reagents (Blake, M.S., *et al.*, *Analyt. Biochem. 136*:175-179 (1984)).

Example 5

Expression of the Outer Membrane Protein P2

The mature-P2 and fusion-P2 constructs were used to transform the expression strain BL21 (DE3) Δ*omp*A. The transformation plates were cultured at 30°C. Colonies of both types were isolated from these plates and analyzed. It was found that virtually all transformants contained the desired plasmid DNA.

Various fusion-P2 and mature-P2 containing clones were then analyzed for protein expression. The clones were induced and grown in LB media containing 0.4 % glucose and 118 μM carbenicillin instead of ampicillin with an aeration speed of 100 to 150 rpm at about 30°C. The expression of the P2 protein was analyzed by loading 0.1 ml of the culture of total *E. coli* proteins on an 8-16% gradient SDS gel (*see* Figs. 1 and 2).

Example 6

Purification and Refolding of the Outer Membrane Protein P2

E. coli strain BL21 (DE3) Δ*omp*A [pNV-3] was grown to mid-log phase (OD = 0.6 at 600 nm) in Luria broth. Isopropyl thiogalactoside was then added (0.4 mM final) and the cells grown an additional three hours at 30°C. The cells were then harvested and washed with several volumes of TEN buffer (50 mM Tris-HCl, 0.2 M NaCl, 10 mM EDTA, at pH 8.0) and the cell paste stored frozen at -75°C.

For purification, about 3 grams of cells were thawed and suspended in 9 ml of TEN buffer. Lysozyme (Sigma, 0.25 mg/ml), then deoxycholate (Sigma, 1.3 mg/ml) plus PMSF (Sigma, 10 μg/ml) were added and the mixture gently shaken for one hour at room temperature. During this time, the cells lysed and the released DNA caused the solution to become very viscous. DNase was then added (Sigma, 2 μg/ml) and the solution again mixed for one hour at room temperature. The mixture was then centrifuged at 15 K rpm in an SA-600 rotor for 30 minutes and the supernatant discarded. The pellet was then twice suspended in 10 ml of TEN buffer and the supernatant discarded. The pellet was then suspended in 10 ml of 8 M urea (Pierce) in TEN buffer. Alternatively, the pellet was suspended in 10 ml of 6 M guanidine HCl (Sigma) in TEN buffer. The mixture was gently stirred to break up any clumps. The suspension was sonicated for 20 minutes or until an even suspension was achieved. Ten ml of a 10% aqueous solution of 3,14-Zwittergent was added and the solution thoroughly mixed. The solution was again sonicated for 10 minutes. Any residual insoluble material was removed by centrifugation.

This mixture was then applied to a 180 x 2.5 cm column of Sephacryl-300 (Pharmacia) equilibrated in 100 mM Tris-HCl, 1 M NaCl, 10 mM EDTA, 20 mM $CaCl_2$ and 0.05% 3,14-Zwittergent, at pH 8.0. The flow rate was maintained at 1 ml/min. Fractions of 10 ml were collected. The porin refolded into trimer during the gel filtration. The $OD_{280\ nm}$ of each fraction was measured and those fractions containing protein were subjected to SDS gel electrophoresis assay for porin. Those fractions containing porin were pooled and stored at 4°C for three weeks. During the incubation at 4°C, a slow conformational change occurred. This was necessary for the protein to remain in solution without the elevated levels of salt. The pooled fractions were then dialyzed against 50 mM Tris-HCl, 200 mM NaCl, 10 mM EDTA and 0.05% 3,14-Zwittergent, at pH 8.0. This material was then applied to a 2.5 x cm Fast Flow Q (Pharmacia) column equilibrated in the same buffer. Any unbound protein was then eluted with starting buffer. A linear 0.2 to 2.0 M

NaCl gradient was then applied to the column. The porin eluted just before the center of the gradient. Fractions were assayed by SDS-PAGE and the purest fractions pooled and dialyzed against TEN buffer containing 0.05% 3,14-Zwittergent.

Example 7

Coupling of Oxidized Hib Capsular Polysaccharide to the Native Haemophilus Influenzae P2 Protein

The oxidized Hib capsular polysaccharide (10.4 mg) was added to native Hib P2 protein (3.1 mg) purified by the method of Munson *et al., J. Clin. Investig. 72*:677-684 (1983), dissolved in 0.21 ml of 0.2 M phosphate buffer, pH 7.5, containing 5% octyl glucoside. After the polysaccharide was dissolved, sodium cyanoborohydride (7 mg) was added and the reaction solution was incubated at 37°C for 4 days. The reaction mixture was then diluted with 0.15 M sodium chloride solution containing 0.01% thimerosal and separated by gel filtration column chromatography using Biogel A-1.5m (Bio-Rad).

The conjugate (Hib-PP) was obtained as a single peak eluting near the void volume. Analysis of the conjugate solution for sialic acid and protein showed that the conjugate consists of 43% polysaccharide by weight. The P2 protein was recovered in the conjugate in 44% yield and the polysaccharide in 12% yield. The protein recoveries in different experiments generally occur in the 50-80% range and those of the polysaccharide in the 9-13% range.

Example 8

Immunogenicity Studies Using Native Hib-PP Conjugate

Immunogenicity studies were performed as follows. The immunogenicities of the Hib-PP conjugate and the Hib Tetanus toxoid (Hib-TT) conjugate, prepared by a similar coupling procedure, were assayed in 7-

week-old New Zealand white rabbits. The polysaccharide conjugates (10 μg) were administered on days 0, 7 and 14, and the sera collected on day 28. The conjugates were administered in saline solutions.

The sera ELISA titers against the polysaccharide antigen are summarized in Table 1, below. "PP" in Table 1 represents the outer membrane porin protein, P2, purified from *Haemophilus influenzae* type b.

Table 1. ELISA Titers of *Haemophilus Influenzae* Type b Conjugate Vaccines (Hib-Protein)

Vaccine	Adjuvant	ELISA Titer
Hib-TT	Saline	270
Hib-PP	Saline	6205
Hib-PP / PP (30 μg)	Saline	8055
Hib, oxidized	Saline	0

Western blot analysis was performed, on both purified recombinant P2 and lysates derived from *E. coli* expressing the recombinant P2, using a polyclonal antisera generated by immunization of rabbits with a conjugate vaccine composed of Hib polysaccharide linked to the native P2 protein isolated from Hib strain A2 which is equivalent to strain Eagan. The antisera used in the Western blot had been previously shown by ELISA analysis to have a anti-polysaccharide titer of 8500 and an anti-P2 titer of 60,000.

Figure 8 shows the results, demonstrating that the polyclonal antisera generated by immunization of rabbits with a conjugate vaccine containing native P2 derived from the Hib bacteria reacted well with the recombinant P2 on a Western blot. This demonstrates the presence of shared epitopes between the native and recombinant P2 proteins.

The recombinant P2 purified from the high expression *E. coli* system resembles native P2 purified from *Haemophilus influenzae* type b organism in the following aspects. First, antibody against native P2 from *H. influenza* reacted well with the recombinant P2 from the high expression *E. coli* system

on a Western blot indicating the presence of shared epitopes between the native and recombinant P2 proteins.

Second, P2 is a porin. Like porins from other gram-negative bacteria, P2 is made up of three identical polypeptide chains and, in their native trimer conformation, form water-filled, voltage-dependent, channels within the outer membrane of the bacteria. The purified recombinant P2 is a trimer as shown in gel filtration chromatography using Superose 12 (Pharmacia). Recombinant P2 eluted from the column corresponding to a molecular weight of 120 kDa. Native P2 from *H. influenzae* and other bacterial porins such as *Neisseria meningitidis* class 2 and 3 porins also eluted in a similar profile. Unfolded P2 is not soluble and elutes from the size column as a monomer. Addition of $CaCl_2$ helps the refolding of P2 into trimeric conformation as shown in Figure 3 herein.

* * * * *

All publications mentioned hereinabove are hereby incorporated in their entirety by reference.

While the foregoing invention has been described in some detail for purposes of clarity and understanding, it will be appreciated by one skilled in the art from a reading of this disclosure that various changes in form and detail can be made without departing from the true scope of the invention and appended claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: North American Vaccine, Inc.
12103 Indian Creek Court
Beltsville, MD 20705

INVENTORS: Tai, Joseph Y.
Pullen, Jeffrey K.
Soper, Thomas S.
Liang, Shu-Mei

(ii) TITLE OF INVENTION: A Method For The High Level Expression, Purification And Refolding Of The Outer Membrane Protein P2 From Haemophilus Influenzae Type b

(iii) NUMBER OF SEQUENCES: 14

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Sterne, Kessler, Goldstein & Fox
(B) STREET: 1100 New York Avenue, Suite 600
(C) CITY: Washington
(D) STATE: D.C.
(E) COUNTRY: U.S.A.
(F) ZIP: 20005-3934

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: (To Be Assigned)
(B) FILING DATE: Herewith
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/096,181
(B) FILING DATE: 23-JULY-1993

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Esmond, Robert W.
(B) REFERENCE/DOCKET NUMBER: 1438.001PC01

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (202) 371-2600
(B) TELEFAX: (202) 371-2540

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

TTCTGGCGAG TCGACAATTC TATTGG 26

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

AACCTTTATC GTCGACGAGC AATTGG 26

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 44 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GCTTCAGCAG CACATATGGC TGTTGTTTAT AACAACGAAG GGAC 44

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 45 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GCAGCTTCAG CAGCGGATCC AGCTGTTGTT TATAACAACG AAGGG 45

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 35 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GCAAAAAAAG CGAATCTCTC GAGTCGCCTT GCTTT 35

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 41 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

AAAAAAAGCG AATCTTTGGA TCCGCCTTGC TTTTAATAAT G 41

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1477 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 65..1147

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```
GTCGACAATT CTATTGGAGA AAAGTTCAAT CATAGATAGT AAACAACCAT AAGGAATACA      60

AATT ATG AAA AAA ACA CTT GCA GCA TTA ATC GTT GGT GCA TTC GCA GCT      109
     Met Lys Lys Thr Leu Ala Ala Leu Ile Val Gly Ala Phe Ala Ala
       1               5                  10                  15

TCA GCA GCA AAC GCA GCT GTT GTT TAT AAC AAC GAA GGG ACT AAC GTA      157
Ser Ala Ala Asn Ala Ala Val Val Tyr Asn Asn Glu Gly Thr Asn Val
                  20                  25                  30

GAA TTA GGT GGT CGT TTA AGC ATT ATC GCA GAA CAA AGT AAT AGC ACT      205
Glu Leu Gly Gly Arg Leu Ser Ile Ile Ala Glu Gln Ser Asn Ser Thr
              35                  40                  45

GTA GAT AAT CAA AAA CAG CAA CAC GGT GCA TTA CGC AAT CAA GGT TCA      253
Val Asp Asn Gln Lys Gln Gln His Gly Ala Leu Arg Asn Gln Gly Ser
          50                  55                  60

CGT TTC CAC ATT AAA GCA ACT CAT AAC TTC GGT GAT GGT TTC TAT GCA      301
Arg Phe His Ile Lys Ala Thr His Asn Phe Gly Asp Gly Phe Tyr Ala
      65                  70                  75

CAA GGT TAT TTA GAA ACT CGT TTT GTT ACA AAA GCC TCT GAA AAC GGT      349
Gln Gly Tyr Leu Glu Thr Arg Phe Val Thr Lys Ala Ser Glu Asn Gly
 80                  85                  90                  95

TCA GAT AAC TTC GGT GAT ATT ACA AGC AAA TAT GCT TAT GTT ACT TTA      397
Ser Asp Asn Phe Gly Asp Ile Thr Ser Lys Tyr Ala Tyr Val Thr Leu
                 100                 105                 110

GGA AAT AAA GCA TTC GGT GAA GTA AAA CTT GGT CGT GCG AAA ACT ATT      445
Gly Asn Lys Ala Phe Gly Glu Val Lys Leu Gly Arg Ala Lys Thr Ile
             115                 120                 125

GCT GAT GGC ATA ACA AGT GCA GAA GAT AAA GAA TAT GGC GTT CTC AAC      493
Ala Asp Gly Ile Thr Ser Ala Glu Asp Lys Glu Tyr Gly Val Leu Asn
         130                 135                 140

AAT AGT GAC TAT ATT CCT ACT AGT GGT AAT ACG GTT GGC TAT ACT TTT      541
Asn Ser Asp Tyr Ile Pro Thr Ser Gly Asn Thr Val Gly Tyr Thr Phe
     145                 150                 155

AAA GGT ATT GAT GGT TTA GTA TTA GGC GCT AAT TAT TTA TTA GCA CAA      589
Lys Gly Ile Asp Gly Leu Val Leu Gly Ala Asn Tyr Leu Leu Ala Gln
160                 165                 170                 175

AAG CGT GAG GGT GCA AAA GGT GAA AAT AAG CGG CCT AAT GAT AAG GCT      637
Lys Arg Glu Gly Ala Lys Gly Glu Asn Lys Arg Pro Asn Asp Lys Ala
                 180                 185                 190

GGT GAA GTA CGT ATA GGT GAA ATC AAT AAT GGA ATT CAA GTT GGT GCA      685
Gly Glu Val Arg Ile Gly Glu Ile Asn Asn Gly Ile Gln Val Gly Ala
             195                 200                 205
```

```
AAA TAT GAT GCA AAC GAC ATC GTT GCA AAA ATT GCT TAT GGT AGA ACT        733
Lys Tyr Asp Ala Asn Asp Ile Val Ala Lys Ile Ala Tyr Gly Arg Thr
        210                 215                 220

AAC TAC AAA TAT AAC GAA TCT GAC GAG CAT AAA CAG CAA TTA AAT GGT        781
Asn Tyr Lys Tyr Asn Glu Ser Asp Glu His Lys Gln Gln Leu Asn Gly
    225                 230                 235

GTA TTA GCA ACT TTA GGC TAT CGT TTT AGT GAT TTA GGC TTA TTA GTG        829
Val Leu Ala Thr Leu Gly Tyr Arg Phe Ser Asp Leu Gly Leu Leu Val
240                 245                 250                 255

TCT CTA GAT AGT GGC TAT GCA AAA ACT AAA AAC TAT AAA ATT AAA CAC        877
Ser Leu Asp Ser Gly Tyr Ala Lys Thr Lys Asn Tyr Lys Ile Lys His
                260                 265                 270

GAA AAA CGC TAT TTC GTA TCT CCA GGT TTC CAA TAT GAA TTA ATG GAA        925
Glu Lys Arg Tyr Phe Val Ser Pro Gly Phe Gln Tyr Glu Leu Met Glu
            275                 280                 285

GAT ACT AAT GTC TAT GGC AAC TTC AAA TAT GAA CGC ACT TCT GTA GAT        973
Asp Thr Asn Val Tyr Gly Asn Phe Lys Tyr Glu Arg Thr Ser Val Asp
        290                 295                 300

CAA GGT GAA AAA ACA CGT GAA CAA GCA GTA TTA TTC GGT GTA GAT CAT       1021
Gln Gly Glu Lys Thr Arg Glu Gln Ala Val Leu Phe Gly Val Asp His
    305                 310                 315

AAA CTT CAC AAA CAA CTA TTA ACC TAT ATT GAA GGT GCT TAC GCT AGA       1069
Lys Leu His Lys Gln Leu Leu Thr Tyr Ile Glu Gly Ala Tyr Ala Arg
320                 325                 330                 335

ACT AGA ACA ACT GAG ACA GGT AAA GGC GTA AAA ACT GAA AAA GAA AAA       1117
Thr Arg Thr Thr Glu Thr Gly Lys Gly Val Lys Thr Glu Lys Glu Lys
                340                 345                 350

TCA GTG GGT GTA GGT TTA CGC GTT TAC TTC TAATCATTTG TTAGAAATAC         1167
Ser Val Gly Val Gly Leu Arg Val Tyr Phe
            355                 360

ATTATTAAAA GCAAGGCGAA TCGAAAGATT CGCTTTTTTT GCTCAAAATC AAGTTAAAAA     1227

ATGATTAAGT TAAAAGTGTA TAAATATTTA GGCTATTTTA TAAGTAACAA AATATTAATA     1287

AAAAATCTGT GACATATATC ACAGATTTTT AAATCAATTA ACTATTTAAG TGTTTACTAT     1347

TAATTCTCTT TCCACTTTCC GTTTACTACT GTGCCGATTA CTTGGTAATT TGGCGTAAAC     1407

ACGGCTAAGT TTGCTATCTT ACCTTTTTCT ACCGAACCTA AACGATCATC TATACCAATT     1467

GCTCGTCGAC                                                           1477
```

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 361 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```
Met Lys Lys Thr Leu Ala Ala Leu Ile Val Gly Ala Phe Ala Ala Ser
  1               5                  10                  15
```

```
Ala Ala Asn Ala Ala Val Val Tyr Asn Asn Glu Gly Thr Asn Val Glu
            20                  25                  30

Leu Gly Gly Arg Leu Ser Ile Ile Ala Glu Gln Ser Asn Ser Thr Val
        35                  40                  45

Asp Asn Gln Lys Gln Gln His Gly Ala Leu Arg Asn Gln Gly Ser Arg
    50                  55                  60

Phe His Ile Lys Ala Thr His Asn Phe Gly Asp Gly Phe Tyr Ala Gln
65                  70                  75                  80

Gly Tyr Leu Glu Thr Arg Phe Val Thr Lys Ala Ser Glu Asn Gly Ser
                85                  90                  95

Asp Asn Phe Gly Asp Ile Thr Ser Lys Tyr Ala Tyr Val Thr Leu Gly
            100                 105                 110

Asn Lys Ala Phe Gly Glu Val Lys Leu Gly Arg Ala Lys Thr Ile Ala
        115                 120                 125

Asp Gly Ile Thr Ser Ala Glu Asp Lys Glu Tyr Gly Val Leu Asn Asn
    130                 135                 140

Ser Asp Tyr Ile Pro Thr Ser Gly Asn Thr Val Gly Tyr Thr Phe Lys
145                 150                 155                 160

Gly Ile Asp Gly Leu Val Leu Gly Ala Asn Tyr Leu Leu Ala Gln Lys
                165                 170                 175

Arg Glu Gly Ala Lys Gly Glu Asn Lys Arg Pro Asn Asp Lys Ala Gly
            180                 185                 190

Glu Val Arg Ile Gly Glu Ile Asn Asn Gly Ile Gln Val Gly Ala Lys
        195                 200                 205

Tyr Asp Ala Asn Asp Ile Val Ala Lys Ile Ala Tyr Gly Arg Thr Asn
    210                 215                 220

Tyr Lys Tyr Asn Glu Ser Asp Glu His Lys Gln Gln Leu Asn Gly Val
225                 230                 235                 240

Leu Ala Thr Leu Gly Tyr Arg Phe Ser Asp Leu Gly Leu Leu Val Ser
                245                 250                 255

Leu Asp Ser Gly Tyr Ala Lys Thr Lys Asn Tyr Lys Ile Lys His Glu
            260                 265                 270

Lys Arg Tyr Phe Val Ser Pro Gly Phe Gln Tyr Glu Leu Met Glu Asp
        275                 280                 285

Thr Asn Val Tyr Gly Asn Phe Lys Tyr Glu Arg Thr Ser Val Asp Gln
    290                 295                 300

Gly Glu Lys Thr Arg Glu Gln Ala Val Leu Phe Gly Val Asp His Lys
305                 310                 315                 320

Leu His Lys Gln Leu Leu Thr Tyr Ile Glu Gly Ala Tyr Ala Arg Thr
                325                 330                 335

Arg Thr Thr Glu Thr Gly Lys Gly Val Lys Thr Glu Lys Glu Lys Ser
            340                 345                 350

Val Gly Val Gly Leu Arg Val Tyr Phe
        355                 360
```

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1137 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 4..1092

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

```
CAT ATG GCT AGC ATG ACT GGT GGA CAG CAA ATG GGT CGG GAT TCA AGC      48
    Met Ala Ser Met Thr Gly Gly Gln Gln Met Gly Arg Asp Ser Ser
      1               5                  10                  15

TTG GTA CCG AGC TCG GAT CCA GCT GTT GTT TAT AAC AAC GAA GGG ACT      96
Leu Val Pro Ser Ser Asp Pro Ala Val Val Tyr Asn Asn Glu Gly Thr
                  20                  25                  30

AAC GTA GAA TTA GGT GGT CGT TTA AGC ATT ATC GCA GAA CAA AGT AAT     144
Asn Val Glu Leu Gly Gly Arg Leu Ser Ile Ile Ala Glu Gln Ser Asn
              35                  40                  45

AGC ACT GTA GAT AAT CAA AAA CAG CAA CAC GGT GCA TTA CGC AAT CAA     192
Ser Thr Val Asp Asn Gln Lys Gln Gln His Gly Ala Leu Arg Asn Gln
          50                  55                  60

GGT TCA CGT TTC CAC ATT AAA GCA ACT CAT AAC TTC GGT GAT GGT TTC     240
Gly Ser Arg Phe His Ile Lys Ala Thr His Asn Phe Gly Asp Gly Phe
      65                  70                  75

TAT GCA CAA GGT TAT TTA GAA ACT CGT TTT GTT ACA AAA GCC TCT GAA     288
Tyr Ala Gln Gly Tyr Leu Glu Thr Arg Phe Val Thr Lys Ala Ser Glu
 80                  85                  90                  95

AAC GGT TCA GAT AAC TTC GGT GAT ATT ACA AGC AAA TAT GCT TAT GTT     336
Asn Gly Ser Asp Asn Phe Gly Asp Ile Thr Ser Lys Tyr Ala Tyr Val
                 100                 105                 110

ACT TTA GGA AAT AAA GCA TTC GGT GAA GTA AAA CTT GGT CGT GCG AAA     384
Thr Leu Gly Asn Lys Ala Phe Gly Glu Val Lys Leu Gly Arg Ala Lys
             115                 120                 125

ACT ATT GCT GAT GGC ATA ACA AGT GCA GAA GAT AAA GAA TAT GGC GTT     432
Thr Ile Ala Asp Gly Ile Thr Ser Ala Glu Asp Lys Glu Tyr Gly Val
         130                 135                 140

CTC AAC AAT AGT GAC TAT ATT CCT ACT AGT GGT AAT ACG GTT GGC TAT     480
Leu Asn Asn Ser Asp Tyr Ile Pro Thr Ser Gly Asn Thr Val Gly Tyr
     145                 150                 155

ACT TTT AAA GGT ATT GAT GGT TTA GTA TTA GGC GCT AAT TAT TTA TTA     528
Thr Phe Lys Gly Ile Asp Gly Leu Val Leu Gly Ala Asn Tyr Leu Leu
160                 165                 170                 175

GCA CAA AAG CGT GAG GGT GCA AAA GGT GAA AAT AAG CGG CCT AAT GAT     576
Ala Gln Lys Arg Glu Gly Ala Lys Gly Glu Asn Lys Arg Pro Asn Asp
                 180                 185                 190

AAG GCT GGT GAA GTA CGT ATA GGT GAA ATC AAT AAT GGA ATT CAA GTT     624
Lys Ala Gly Glu Val Arg Ile Gly Glu Ile Asn Asn Gly Ile Gln Val
             195                 200                 205
```

```
GGT GCA AAA TAT GAT GCA AAC GAC ATC GTT GCA AAA ATT GCT TAT GGT      672
Gly Ala Lys Tyr Asp Ala Asn Asp Ile Val Ala Lys Ile Ala Tyr Gly
        210                 215                 220

AGA ACT AAC TAC AAA TAT AAC GAA TCT GAC GAG CAT AAA CAG CAA TTA      720
Arg Thr Asn Tyr Lys Tyr Asn Glu Ser Asp Glu His Lys Gln Gln Leu
    225                 230                 235

AAT GGT GTA TTA GCA ACT TTA GGC TAT CGT TTT AGT GAT TTA GGC TTA      768
Asn Gly Val Leu Ala Thr Leu Gly Tyr Arg Phe Ser Asp Leu Gly Leu
240                 245                 250                 255

TTA GTG TCT CTA GAT AGT GGC TAT GCA AAA ACT AAA AAC TAT AAA ATT      816
Leu Val Ser Leu Asp Ser Gly Tyr Ala Lys Thr Lys Asn Tyr Lys Ile
                260                 265                 270

AAA CAC GAA AAA CGC TAT TTC GTA TCT CCA GGT TTC CAA TAT GAA TTA      864
Lys His Glu Lys Arg Tyr Phe Val Ser Pro Gly Phe Gln Tyr Glu Leu
            275                 280                 285

ATG GAA GAT ACT AAT GTC TAT GGC AAC TTC AAA TAT GAA CGC ACT TCT      912
Met Glu Asp Thr Asn Val Tyr Gly Asn Phe Lys Tyr Glu Arg Thr Ser
        290                 295                 300

GTA GAT CAA GGT GAA AAA ACA CGT GAA CAA GCA GTA TTA TTC GGT GTA      960
Val Asp Gln Gly Glu Lys Thr Arg Glu Gln Ala Val Leu Phe Gly Val
    305                 310                 315

GAT CAT AAA CTT CAC AAA CAA CTA TTA ACC TAT ATT GAA GGT GCT TAC     1008
Asp His Lys Leu His Lys Gln Leu Leu Thr Tyr Ile Glu Gly Ala Tyr
320                 325                 330                 335

GCT AGA ACT AGA ACA ACT GAG ACA GGT AAA GGC GTA AAA ACT GAA AAA     1056
Ala Arg Thr Arg Thr Thr Glu Thr Gly Lys Gly Val Lys Thr Glu Lys
                340                 345                 350

GAA AAA TCA GTG GGT GTA GGT TTA CGC GTT TAC TTC TAATCATTTG          1102
Glu Lys Ser Val Gly Val Gly Leu Arg Val Tyr Phe
            355                 360

TTAGAAATAC ATTATTAAAA GCAAGGCGAC TCGAG                              1137
```

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 363 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

```
Met Ala Ser Met Thr Gly Gly Gln Gln Met Gly Arg Asp Ser Ser Leu
  1               5                  10                  15

Val Pro Ser Ser Asp Pro Ala Val Val Tyr Asn Asn Glu Gly Thr Asn
            20                  25                  30

Val Glu Leu Gly Gly Arg Leu Ser Ile Ile Ala Glu Gln Ser Asn Ser
        35                  40                  45

Thr Val Asp Asn Gln Lys Gln Gln His Gly Ala Leu Arg Asn Gln Gly
    50                  55                  60
```

```
Ser Arg Phe His Ile Lys Ala Thr His Asn Phe Gly Asp Gly Phe Tyr
 65                  70                  75                  80

Ala Gln Gly Tyr Leu Glu Thr Arg Phe Val Thr Lys Ala Ser Glu Asn
                 85                  90                  95

Gly Ser Asp Asn Phe Gly Asp Ile Thr Ser Lys Tyr Ala Tyr Val Thr
            100                 105                 110

Leu Gly Asn Lys Ala Phe Gly Glu Val Lys Leu Gly Arg Ala Lys Thr
        115                 120                 125

Ile Ala Asp Gly Ile Thr Ser Ala Glu Asp Lys Glu Tyr Gly Val Leu
    130                 135                 140

Asn Asn Ser Asp Tyr Ile Pro Thr Ser Gly Asn Thr Val Gly Tyr Thr
145                 150                 155                 160

Phe Lys Gly Ile Asp Gly Leu Val Leu Gly Ala Asn Tyr Leu Leu Ala
                165                 170                 175

Gln Lys Arg Glu Gly Ala Lys Gly Glu Asn Lys Arg Pro Asn Asp Lys
            180                 185                 190

Ala Gly Glu Val Arg Ile Gly Glu Ile Asn Asn Gly Ile Gln Val Gly
        195                 200                 205

Ala Lys Tyr Asp Ala Asn Asp Ile Val Ala Lys Ile Ala Tyr Gly Arg
    210                 215                 220

Thr Asn Tyr Lys Tyr Asn Glu Ser Asp Glu His Lys Gln Gln Leu Asn
225                 230                 235                 240

Gly Val Leu Ala Thr Leu Gly Tyr Arg Phe Ser Asp Leu Gly Leu Leu
                245                 250                 255

Val Ser Leu Asp Ser Gly Tyr Ala Lys Thr Lys Asn Tyr Lys Ile Lys
            260                 265                 270

His Glu Lys Arg Tyr Phe Val Ser Pro Gly Phe Gln Tyr Glu Leu Met
        275                 280                 285

Glu Asp Thr Asn Val Tyr Gly Asn Phe Lys Tyr Glu Arg Thr Ser Val
    290                 295                 300

Asp Gln Gly Glu Lys Thr Arg Glu Gln Ala Val Leu Phe Gly Val Asp
305                 310                 315                 320

His Lys Leu His Lys Gln Leu Leu Thr Tyr Ile Glu Gly Ala Tyr Ala
                325                 330                 335

Arg Thr Arg Thr Thr Glu Thr Gly Lys Gly Val Lys Thr Glu Lys Glu
            340                 345                 350

Lys Ser Val Gly Val Gly Leu Arg Val Tyr Phe
        355                 360
```

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1074 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 4..1029

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

```
CAT ATG GCT GTT GTT TAT AAC AAC GAA GGG ACT AAC GTA GAA TTA GGT      48
    Met Ala Val Val Tyr Asn Asn Glu Gly Thr Asn Val Glu Leu Gly
      1               5                  10                  15

GGT CGT TTA AGC ATT ATC GCA GAA CAA AGT AAT AGC ACT GTA GAT AAT      96
Gly Arg Leu Ser Ile Ile Ala Glu Gln Ser Asn Ser Thr Val Asp Asn
                20                  25                  30

CAA AAA CAG CAA CAC GGT GCA TTA CGC AAT CAA GGT TCA CGT TTC CAC     144
Gln Lys Gln Gln His Gly Ala Leu Arg Asn Gln Gly Ser Arg Phe His
            35                  40                  45

ATT AAA GCA ACT CAT AAC TTC GGT GAT GGT TTC TAT GCA CAA GGT TAT     192
Ile Lys Ala Thr His Asn Phe Gly Asp Gly Phe Tyr Ala Gln Gly Tyr
        50                  55                  60

TTA GAA ACT CGT TTT GTT ACA AAA GCC TCT GAA AAC GGT TCA GAT AAC     240
Leu Glu Thr Arg Phe Val Thr Lys Ala Ser Glu Asn Gly Ser Asp Asn
    65                  70                  75

TTC GGT GAT ATT ACA AGC AAA TAT GCT TAT GTT ACT TTA GGA AAT AAA     288
Phe Gly Asp Ile Thr Ser Lys Tyr Ala Tyr Val Thr Leu Gly Asn Lys
80                  85                  90                  95

GCA TTC GGT GAA GTA AAA CTT GGT CGT GCG AAA ACT ATT GCT GAT GGC     336
Ala Phe Gly Glu Val Lys Leu Gly Arg Ala Lys Thr Ile Ala Asp Gly
                100                 105                 110

ATA ACA AGT GCA GAA GAT AAA GAA TAT GGC GTT CTC AAC AAT AGT GAC     384
Ile Thr Ser Ala Glu Asp Lys Glu Tyr Gly Val Leu Asn Asn Ser Asp
            115                 120                 125

TAT ATT CCT ACT AGT GGT AAT ACG GTT GGC TAT ACT TTT AAA GGT ATT     432
Tyr Ile Pro Thr Ser Gly Asn Thr Val Gly Tyr Thr Phe Lys Gly Ile
        130                 135                 140

GAT GGT TTA GTA TTA GGC GCT AAT TAT TTA TTA GCA CAA AAG CGT GAG     480
Asp Gly Leu Val Leu Gly Ala Asn Tyr Leu Leu Ala Gln Lys Arg Glu
    145                 150                 155

GGT GCA AAA GGT GAA AAT AAG CGG CCT AAT GAT AAG GCT GGT GAA GTA     528
Gly Ala Lys Gly Glu Asn Lys Arg Pro Asn Asp Lys Ala Gly Glu Val
160                 165                 170                 175

CGT ATA GGT GAA ATC AAT AAT GGA ATT CAA GTT GGT GCA AAA TAT GAT     576
Arg Ile Gly Glu Ile Asn Asn Gly Ile Gln Val Gly Ala Lys Tyr Asp
                180                 185                 190

GCA AAC GAC ATC GTT GCA AAA ATT GCT TAT GGT AGA ACT AAC TAC AAA     624
Ala Asn Asp Ile Val Ala Lys Ile Ala Tyr Gly Arg Thr Asn Tyr Lys
            195                 200                 205

TAT AAC GAA TCT GAC GAG CAT AAA CAG CAA TTA AAT GGT GTA TTA GCA     672
Tyr Asn Glu Ser Asp Glu His Lys Gln Gln Leu Asn Gly Val Leu Ala
        210                 215                 220

ACT TTA GGC TAT CGT TTT AGT GAT TTA GGC TTA TTA GTG TCT CTA GAT     720
Thr Leu Gly Tyr Arg Phe Ser Asp Leu Gly Leu Leu Val Ser Leu Asp
    225                 230                 235
```

```
AGT GGC TAT GCA AAA ACT AAA AAC TAT AAA ATT AAA CAC GAA AAA CGC      768
Ser Gly Tyr Ala Lys Thr Lys Asn Tyr Lys Ile Lys His Glu Lys Arg
240                 245                 250                 255

TAT TTC GTA TCT CCA GGT TTC CAA TAT GAA TTA ATG GAA GAT ACT AAT      816
Tyr Phe Val Ser Pro Gly Phe Gln Tyr Glu Leu Met Glu Asp Thr Asn
                260                 265                 270

GTC TAT GGC AAC TTC AAA TAT GAA CGC ACT TCT GTA GAT CAA GGT GAA      864
Val Tyr Gly Asn Phe Lys Tyr Glu Arg Thr Ser Val Asp Gln Gly Glu
            275                 280                 285

AAA ACA CGT GAA CAA GCA GTA TTA TTC GGT GTA GAT CAT AAA CTT CAC      912
Lys Thr Arg Glu Gln Ala Val Leu Phe Gly Val Asp His Lys Leu His
        290                 295                 300

AAA CAA CTA TTA ACC TAT ATT GAA GGT GCT TAC GCT AGA ACT AGA ACA      960
Lys Gln Leu Leu Thr Tyr Ile Glu Gly Ala Tyr Ala Arg Thr Arg Thr
    305                 310                 315

ACA GAG ACA GGT AAA GGC GTA AAA ACT GAA AAA GAA AAA TCA GTG GGT     1008
Thr Glu Thr Gly Lys Gly Val Lys Thr Glu Lys Glu Lys Ser Val Gly
320                 325                 330                 335

GTA GGT TTA CGC GTT TAC TTC TAATCATTTG TTAGAAATAC ATTATTAAAA        1059
Val Gly Leu Arg Val Tyr Phe
                340

GCAAGGCGAC TCGAG                                                   1074
```

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 342 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

```
Met Ala Val Val Tyr Asn Asn Glu Gly Thr Asn Val Glu Leu Gly Gly
  1               5                  10                  15

Arg Leu Ser Ile Ile Ala Glu Gln Ser Asn Ser Thr Val Asp Asn Gln
            20                  25                  30

Lys Gln Gln His Gly Ala Leu Arg Asn Gln Gly Ser Arg Phe His Ile
        35                  40                  45

Lys Ala Thr His Asn Phe Gly Asp Gly Phe Tyr Ala Gln Gly Tyr Leu
    50                  55                  60

Glu Thr Arg Phe Val Thr Lys Ala Ser Glu Asn Gly Ser Asp Asn Phe
65                  70                  75                  80

Gly Asp Ile Thr Ser Lys Tyr Ala Tyr Val Thr Leu Gly Asn Lys Ala
                85                  90                  95

Phe Gly Glu Val Lys Leu Gly Arg Ala Lys Thr Ile Ala Asp Gly Ile
            100                 105                 110

Thr Ser Ala Glu Asp Lys Glu Tyr Gly Val Leu Asn Asn Ser Asp Tyr
        115                 120                 125
```

```
Ile Pro Thr Ser Gly Asn Thr Val Gly Tyr Thr Phe Lys Gly Ile Asp
    130                 135                 140

Gly Leu Val Leu Gly Ala Asn Tyr Leu Leu Ala Gln Lys Arg Glu Gly
145                 150                 155                 160

Ala Lys Gly Glu Asn Lys Arg Pro Asn Asp Lys Ala Gly Glu Val Arg
                165                 170                 175

Ile Gly Glu Ile Asn Asn Gly Ile Gln Val Gly Ala Lys Tyr Asp Ala
            180                 185                 190

Asn Asp Ile Val Ala Lys Ile Ala Tyr Gly Arg Thr Asn Tyr Lys Tyr
        195                 200                 205

Asn Glu Ser Asp Glu His Lys Gln Gln Leu Asn Gly Val Leu Ala Thr
    210                 215                 220

Leu Gly Tyr Arg Phe Ser Asp Leu Gly Leu Leu Val Ser Leu Asp Ser
225                 230                 235                 240

Gly Tyr Ala Lys Thr Lys Asn Tyr Lys Ile Lys His Glu Lys Arg Tyr
                245                 250                 255

Phe Val Ser Pro Gly Phe Gln Tyr Glu Leu Met Glu Asp Thr Asn Val
            260                 265                 270

Tyr Gly Asn Phe Lys Tyr Glu Arg Thr Ser Val Asp Gln Gly Glu Lys
        275                 280                 285

Thr Arg Glu Gln Ala Val Leu Phe Gly Val Asp His Lys Leu His Lys
    290                 295                 300

Gln Leu Leu Thr Tyr Ile Glu Gly Ala Tyr Ala Arg Thr Arg Thr Thr
305                 310                 315                 320

Glu Thr Gly Lys Gly Val Lys Thr Glu Lys Glu Lys Ser Val Gly Val
                325                 330                 335

Gly Leu Arg Val Tyr Phe
            340
```

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1072 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 4..1029

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

```
CAT ATG GCT GTT GTT TAT AAC AAC GAA GGG ACT AAC GTA GAA TTA GGT      48
    Met Ala Val Val Tyr Asn Asn Glu Gly Thr Asn Val Glu Leu Gly
      1               5                  10                  15

GGT CGT TTA AGC ATT ATC GCA GAA CAA AGT AAT AGC ACT GTA GAT AAT      96
Gly Arg Leu Ser Ile Ile Ala Glu Gln Ser Asn Ser Thr Val Asp Asn
                20                  25                  30
```

```
CAA AAA CAG CAA CAC GGT GCA TTA CGC AAT CAA GGT TCA CGT TTC CAC          144
Gln Lys Gln Gln His Gly Ala Leu Arg Asn Gln Gly Ser Arg Phe His
            35                  40                  45

ATT AAA GCA ACT CAT AAC TTC GGT GAT GGT TTC TAT GCA CAA GGT TAT          192
Ile Lys Ala Thr His Asn Phe Gly Asp Gly Phe Tyr Ala Gln Gly Tyr
        50                  55                  60

TTA GAA ACT CGT TTT GTT ACA AAA GCC TCT GAA AAC GGT TCA GAT AAC          240
Leu Glu Thr Arg Phe Val Thr Lys Ala Ser Glu Asn Gly Ser Asp Asn
    65                  70                  75

TTC GGT GAT ATT ACA AGC AAA TAT GCT TAT GTT ACT TTA GGA AAT AAA          288
Phe Gly Asp Ile Thr Ser Lys Tyr Ala Tyr Val Thr Leu Gly Asn Lys
80                  85                  90                  95

GCA TTC GGT GAA GTA AAA CTT GGT CGT GCG AAA ACT ATT GCT TTT GGC          336
Ala Phe Gly Glu Val Lys Leu Gly Arg Ala Lys Thr Ile Ala Phe Gly
                100                 105                 110

ATA ACA AGT GCA GAA GAT AAA GAA TAT GGC GTT CTC AAC AAT AGT GAC          384
Ile Thr Ser Ala Glu Asp Lys Glu Tyr Gly Val Leu Asn Asn Ser Asp
            115                 120                 125

TAT ATT CCT ACT AGT GGT AAT ACG GTT GGC TAT ACT TTT AAA GGT ATT          432
Tyr Ile Pro Thr Ser Gly Asn Thr Val Gly Tyr Thr Phe Lys Gly Ile
        130                 135                 140

GAT GGT TTA GTA TTA GGC GCT AAT TAT TTA TTA GCA CAA AAG CGT GAG          480
Asp Gly Leu Val Leu Gly Ala Asn Tyr Leu Leu Ala Gln Lys Arg Glu
    145                 150                 155

GGT GCA AAA GGT GAA AAT AAG CGG CCT AAT GAT AAG GCT GGT GAA GTA          528
Gly Ala Lys Gly Glu Asn Lys Arg Pro Asn Asp Lys Ala Gly Glu Val
160                 165                 170                 175

CGT ATA GGT GAA ATC AAT AAT GGA ATT CAA GTT GGT GCA AAA TAT GAT          576
Arg Ile Gly Glu Ile Asn Asn Gly Ile Gln Val Gly Ala Lys Tyr Asp
                180                 185                 190

GCA AAC GAC ATC GTT GCA AAA ATT GCT TAT GGT AGA ACT AAC TAC AAA          624
Ala Asn Asp Ile Val Ala Lys Ile Ala Tyr Gly Arg Thr Asn Tyr Lys
            195                 200                 205

TAT AAC GAA TCT GAC GAG CAT AAA CAG CAA TTA AAT GGT GTA TTA GCA          672
Tyr Asn Glu Ser Asp Glu His Lys Gln Gln Leu Asn Gly Val Leu Ala
        210                 215                 220

ACT TTA GGC TAT CGT TTT AGT GAT TTA GGC TTA TTA GTG TCT CTA GAT          720
Thr Leu Gly Tyr Arg Phe Ser Asp Leu Gly Leu Leu Val Ser Leu Asp
    225                 230                 235

AGT GGC TAT GCA AAA ACT AAA AAC TAT AAA ATT AAA CAC GAA AAA CGC          768
Ser Gly Tyr Ala Lys Thr Lys Asn Tyr Lys Ile Lys His Glu Lys Arg
240                 245                 250                 255

TAT TTC GTA TCT CCA GGT TTC CAA TAT GAA TTA ATG GAA GAT ACT AAT          816
Tyr Phe Val Ser Pro Gly Phe Gln Tyr Glu Leu Met Glu Asp Thr Asn
                260                 265                 270

GTC TAT GGC AAC TTC AAA TAT GAA CGC ACT TCT GTA GAT CAA GGT GAA          864
Val Tyr Gly Asn Phe Lys Tyr Glu Arg Thr Ser Val Asp Gln Gly Glu
            275                 280                 285

AAA ACA CGT GAA CAA GCA GTA TTA TTC GGT GTA GAT CAT AAA CTT CAC          912
Lys Thr Arg Glu Gln Ala Val Leu Phe Gly Val Asp His Lys Leu His
        290                 295                 300
```

```
AAA CAA CTA TTA ACC TAT ATT GAA GGT GCT TAC GCT AGA ACT AGA ACA        960
Lys Gln Leu Leu Thr Tyr Ile Glu Gly Ala Tyr Ala Arg Thr Arg Thr
    305                 310                 315

ACT GAG ACA GGT AAA GGC GTA AAA ACT GAA AAA GAA AAA TCA GTG GGT       1008
Thr Glu Thr Gly Lys Gly Val Lys Thr Glu Lys Glu Lys Ser Val Gly
320                 325                 330                 335

GTA GGT TTA CGC GTT TAC TTC TAATCATTTG TTAGAAATAC ATTATTAAAA          1059
Val Gly Leu Arg Val Tyr Phe
                340

GCAAGGCGGA TCC                                                       1072
```

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 342 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

```
Met Ala Val Val Tyr Asn Asn Glu Gly Thr Asn Val Glu Leu Gly Gly
 1               5                  10                  15

Arg Leu Ser Ile Ile Ala Glu Gln Ser Asn Ser Thr Val Asp Asn Gln
            20                  25                  30

Lys Gln Gln His Gly Ala Leu Arg Asn Gln Gly Ser Arg Phe His Ile
        35                  40                  45

Lys Ala Thr His Asn Phe Gly Asp Gly Phe Tyr Ala Gln Gly Tyr Leu
    50                  55                  60

Glu Thr Arg Phe Val Thr Lys Ala Ser Glu Asn Gly Ser Asp Asn Phe
65                  70                  75                  80

Gly Asp Ile Thr Ser Lys Tyr Ala Tyr Val Thr Leu Gly Asn Lys Ala
                85                  90                  95

Phe Gly Glu Val Lys Leu Gly Arg Ala Lys Thr Ile Ala Asp Gly Ile
            100                 105                 110

Thr Ser Ala Glu Asp Lys Glu Tyr Gly Val Leu Asn Asn Ser Asp Tyr
        115                 120                 125

Ile Pro Thr Ser Gly Asn Thr Val Gly Tyr Thr Phe Lys Gly Ile Asp
    130                 135                 140

Gly Leu Val Leu Gly Ala Asn Tyr Leu Leu Ala Gln Lys Arg Glu Gly
145                 150                 155                 160

Ala Lys Gly Glu Asn Lys Arg Pro Asn Asp Lys Ala Gly Glu Val Arg
                165                 170                 175

Ile Gly Glu Ile Asn Asn Gly Ile Gln Val Gly Ala Lys Tyr Asp Ala
            180                 185                 190

Asn Asp Ile Val Ala Lys Ile Ala Tyr Gly Arg Thr Asn Tyr Lys Tyr
        195                 200                 205
```

```
Asn Glu Ser Asp Glu His Lys Gln Gln Leu Asn Gly Val Leu Ala Thr
    210                 215                 220

Leu Gly Tyr Arg Phe Ser Asp Leu Gly Leu Leu Val Ser Leu Asp Ser
225                 230                 235                 240

Gly Tyr Ala Lys Thr Lys Asn Tyr Lys Ile Lys His Glu Lys Arg Tyr
                245                 250                 255

Phe Val Ser Pro Gly Phe Gln Tyr Glu Leu Met Glu Asp Thr Asn Val
            260                 265                 270

Tyr Gly Asn Phe Lys Tyr Glu Arg Thr Ser Val Asp Gln Gly Glu Lys
        275                 280                 285

Thr Arg Glu Gln Ala Val Leu Phe Gly Val Asp His Lys Leu His Lys
    290                 295                 300

Gln Leu Leu Thr Tyr Ile Glu Gly Ala Tyr Ala Arg Thr Arg Thr Thr
305                 310                 315                 320

Glu Thr Gly Lys Gly Val Lys Thr Glu Lys Glu Lys Ser Val Gly Val
                325                 330                 335

Gly Leu Arg Val Tyr Phe
            340
```

What Is Claimed Is:

1. A method for the high level expression of the outer membrane protein P2 from *Haemophilus influenzae* type b (Hib-P2) in *E. coli* comprising:

(a) transforming a vector comprising a selectable marker and a gene coding for a protein selected from the group consisting of

(i) a mature P2 protein and

(ii) a fusion protein comprising a mature P2 protein fused to amino acids 1 to 22 of the T7 gene ϕ10 capsid protein;

wherein said gene is operably linked to the T7 promoter; and

(b) growing said transformed *E. coli* in LB media containing glucose and a selection agent at about 30°C; whereby the protein is expressed,

wherein the protein so expressed comprises more than about 2% of the total protein expressed in said *E. coli*.

2. The method according to claim 1, wherein said protein comprises more than about 10% of the total protein expressed in said *E. coli*.

3. The method according to claim 1, wherein said protein comprises more than about 40% of the total protein expressed in said *E. coli*.

4. The method according to claim 1, wherein said vector is selected from the group consisting of pET-17b, pET-11a, pET-24a-d(+) and pET-9a.

5. The method according to claim 1, wherein said vector comprises a Hib-P2 gene operably linked to the T7 promoter of expression plasmid pET-17b.

6. A method of purifying the outer membrane protein P2 or fusion protein thereof obtained according to claim 1 comprising:

(c) lysing said *E. coli* obtained in step (b) to release said protein as insoluble inclusion bodies;

(d) washing said insoluble inclusion bodies obtained in step (c) with a buffer to remove contaminating *E. coli* cellular proteins;

(e) suspending and dissolving said inclusion bodies obtained in step (d) in an aqueous solution of a denaturant;

(f) diluting the solution obtained in step (e) with a detergent; and

(g) purifying said protein by gel filtration.

7. A method of refolding the outer membrane protein P2 or fusion protein obtained according to claim 1 comprising:

(c) lysing said *E. coli* obtained in step (b) to release said protein as insoluble inclusion bodies;

(d) washing said insoluble inclusion bodies obtained in step (c) with a buffer to remove contaminating *E. coli* cellular proteins;

(e) suspending and dissolving said inclusion bodies obtained in step (d) in an aqueous solution of a denaturant;

(f) diluting the solution obtained in step (e) with a detergent;

(g) purifying said protein by gel filtration; and

(h) storing said gel filtration product at about 4°C in an aqueous solution comprising high concentration of NaCl and calcium ions, until said protein refolds.

8. A substantially pure refolded outer membrane protein P2 from *Haemophilus influenzae* type b (Hib-P2) or fusion protein thereof produced according to the method of claim 7.

9. A vaccine comprising the outer membrane protein P2 from *Haemophilus influenzae* type b (Hib-P2) or a fusion protein thereof produced according to claim 7 together with a pharmaceutically acceptable diluent, carrier or excipient, wherein said protein is present, in an amount effective to elicit protective antibodies in an animal to *Haemophilus influenzae* type b.

10. The vaccine according to claim 9, wherein said outer membrane protein P2 is conjugated to a *Haemophilus* capsular polysaccharide.

11. A method of obtaining a P2 protein or P2 fusion protein-polysaccharide conjugate comprising:

(i) obtaining the outer membrane protein P2 or fusion protein according to claim 7;

(j) obtaining a *Haemophilus* capsular polysaccharide; and

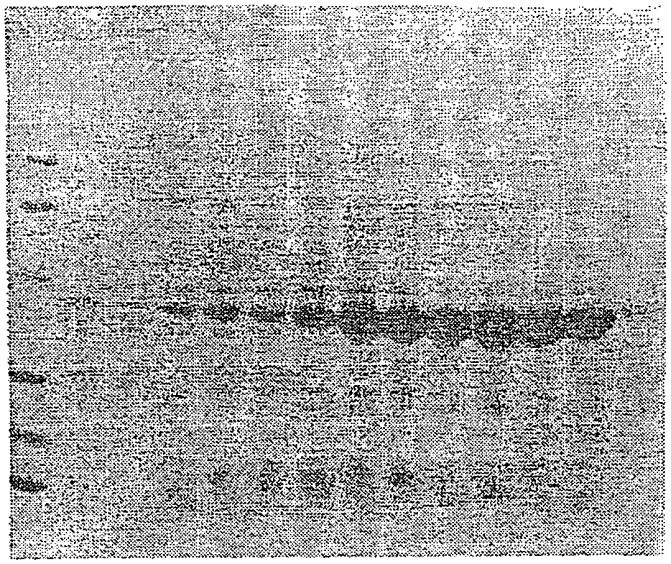
(k) conjugating the outer membrane protein P2 or fusion protein of (i) to the polysaccharide of (j).

12. A method of preventing bacterial meningitis in an animal comprising administering to said animal the Hib-P2 protein or fusion protein produced according to claim 1, wherein said protein is administered in an amount effective to prevent bacterial meningitis.

13. The vector pNV-3.

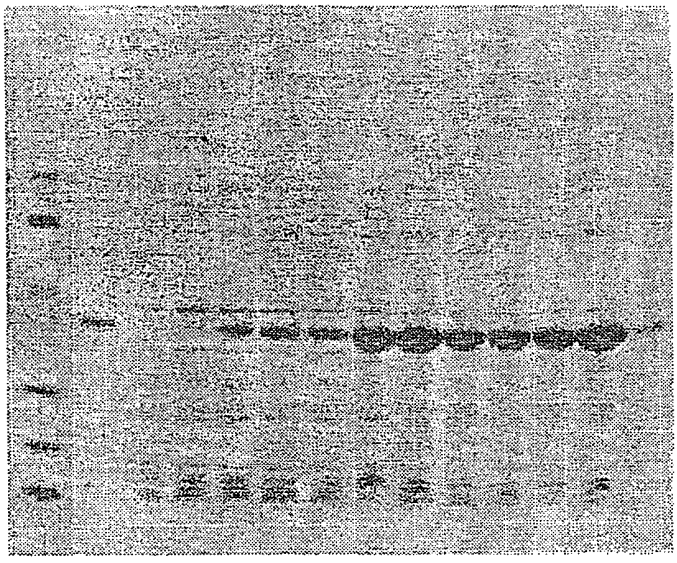
14. The vector pNV-2.

15. The vector pNV-6.



1 2 3 4 5 6 7 8 9 10 11 12 13 14

FIG. 1



1 2 3 4 5 6 7 8 9 10 11 12 13 14

FIG. 2

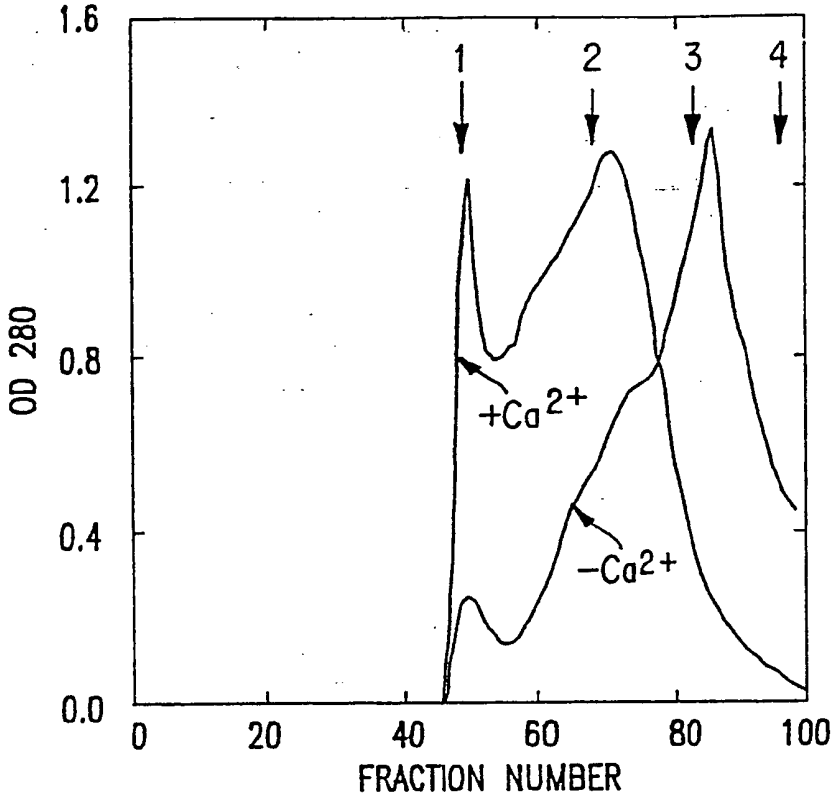

1.6
1.2
0.8
0.4
0.0
OD 280
1
2
3
4
+Ca2+
−Ca2+
0
20
40
60
80
100
FRACTION NUMBER

FIG.3

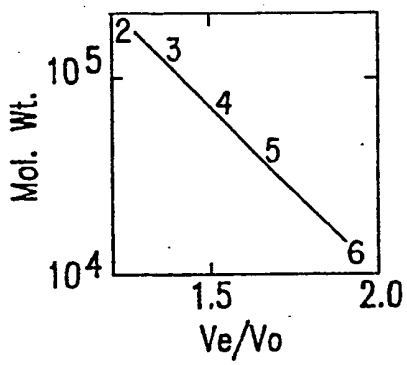

Mol. Wt.
10^5
10^4
2
3
4
5
6
1.5
2.0
Ve/Vo

FIG.3A

SalI → oligo #1 →

GTCGACAATT CTATTGGAGA AAAGTTCAAT CATAGATAGT AAACAACCAT AAGGAATACA 60

AseI

AATT ATG AAA AAA ACA CTT GCA GCA TTA ATC GTT GGT GCA TTC GCA GCT 109
Met Lys Lys Thr Leu Ala Ala Leu Ile Val Gly Ala Phe Ala Ala
1 5 10 15

PvuII → oligo #2 →

TCA GCA GCA AAC GCA GCT GTT GTT TAT AAC AAC GAA GGG ACT AAC GTA 157
Ser Ala Ala Asn Ala Ala Val Val Tyr Asn Asn Glu Gly Thr Asn Val
20 25 30

← oligo #5 ←

GAA TTA GGT GGT CGT TTA AGC ATT ATC GCA GAA CAA AGT AAT AGC ACT 205
Glu Leu Gly Gly Arg Leu Ser Ile Ile Ala Glu Gln Ser Asn Ser Thr
35 40 45

→ oligo #12 →

GTA GAT AAT CAA AAA CAG CAA CAC GGT GCA TTA CGC AAT CAA GGT TCA 253
Val Asp Asn Gln Lys Gln Gln His Gly Ala Leu Arg Asn Gln Gly Ser
50 55 60

← oligo #13 ←

CGT TTC CAC ATT AAA GCA ACT CAT AAC TTC GGT GAT GGT TTC TAT GCA 301
Arg Phe His Ile Lys Ala Thr His Asn Phe Gly Asp Gly Phe Tyr Ala
65 70 75

CAA GGT TAT TTA GAA ACT CGT TTT GTT ACA AAA GCC TCT GAA AAC GGT 349
Gln Gly Tyr Leu Glu Thr Arg Phe Val Thr Lys Ala Ser Glu Asn Gly
80 85 90 95

TCA GAT AAC TTC GGT GAT ATT ACA AGC AAA TAT GCT TAT GTT ACT TTA 397
Ser Asp Asn Phe Gly Asp Ile Thr Ser Lys Tyr Ala Tyr Val Thr Leu
100 105 110

← oligo #6 →

GGA AAT AAA GCA TTC GGT GAA GTA AAA CTT GGT CGT GCG AAA ACT ATT 445
Gly Asn Lys Ala Phe Gly Glu Val Lys Leu Gly Arg Ala Lys Thr Ile
115 120 125

FIG.4A

—oligo #7—

GCT GAT GGC ATA ACA AGT GCA GAA GAT AAA GAA TAT GGC GTT CTC AAC 493
Ala Asp Gly Ile Thr Ser Ala Glu Asp Lys Glu Tyr Gly Val Leu Asn
130 135 140

SpeI DraI

AAT AGT GAC TAT ATT CCT ACT AGT GGT AAT ACG GTT GGC TAT ACT TTT 541
Asn Ser Asp Tyr Ile Pro Thr Ser Gly Asn Thr Val Gly Tyr Thr Phe
145 150 155

HhaI

AAA GGT ATT GAT GGT TTA GTA TTA GGC GCT AAT TAT TTA TTA GCA CAA 589
Lys Gly Ile Asp Gly Leu Val Leu Gly Ala Asn Tyr Leu Leu Ala Gln
160 165 170 175

—oligo #8—

FnuDI

AAG CGT GAG GGT GCA AAA GGT GAA AAT AAG CGG CCT AAT GAT AAG GCT 637
Lys Arg Glu Gly Ala Lys Gly Glu Asn Lys Arg Pro Asn Asp Lys Ala
180 185 190

SnaB1 EcoRI

GGT GAA GTA CGT ATA GGT GAA ATC AAT AAT GGA ATT CAA GTT GGT GCA 685
Gly Glu Val Arg Ile Gly Glu Ile Asn Asn Gly Ile Gln Val Gly Ala
195 200 205

AAA TAT GAT GCA AAC GAC ATC GTT GCA AAA ATT GCT TAT GGT AGA ACT 733
Lys Tyr Asp Ala Asn Asp Ile Val Ala Lys Ile Ala Tyr Gly Arg Thr
210 215 220

—oligo #11—

AAC TAC AAA TAT AAC GAA TCT GAC GAG CAT AAA CAG CAA TTA AAT GGT 781
Asn Tyr Lys Tyr Asn Glu Ser Asp Glu His Lys Gln Gln Leu Asn Gly
225 230 235

GTA TTA GCA ACT TTA GGC TAT CGT TTT AGT GAT TTA GGC TTA TTA GTG 829
Val Leu Ala Thr Leu Gly Tyr Arg Phe Ser Asp Leu Gly Leu Leu Val
240 245 250 255

XbaI —oligo #3— —oligo #14—

TCT CTA GAT AGT GGC TAT GCA AAA ACT AAA AAC TAT AAA ATT AAA CAC 877
Ser Leu Asp Ser Gly Tyr Ala Lys Thr Lys Asn Tyr Lys Ile Lys His
260 265 270

FIG.4B

AseI

GAA AAA CGC TAT TTC GTA TCT CCA GGT TTC CAA TAT GAA TTA ATG GAA 925
Glu Lys Arg Tyr Phe Val Ser Pro Gly Phe Gln Tyr Glu Leu Met Glu
275 280 285

GAT ACT AAT GTC TAT GGC AAC TTC AAA TAT GAA CGC ACT TCT GTA GAT 973
Asp Thr Asn Val Tyr Gly Asn Phe Lys Tyr Glu Arg Thr Ser Val Asp
290 295 300

← oligo #16 → ← oligo #9 — Sau3A

CAA GGT GAA AAA ACA CGT GAA CAA GCA GTA TTA TTC GGT GTA GAT CAT 1021
Gln Gly Glu Lys Thr Arg Glu Gln Ala Val Leu Phe Gly Val Asp His
305 310 315

AAA CTT CAC AAA CAA CTA TTA ACC TAT ATT GAA GGT GCT TAC GCT AGA 1069
Lys Leu His Lys Gln Leu Leu Thr Tyr Ile Glu Gly Ala Tyr Ala Arg
320 325 330 335

ACT AGA ACA ACT GAG ACA GGT AAA GGC GTA AAA ACT GAA AAA GAA AAA 1117
Thr Arg Thr Thr Glu Thr Gly Lys Gly Val Lys Thr Glu Lys Glu Lys
340 345 350

← oligo #15 ← MluI

TCA GTG GGT GTA GGT TTA CGC GTT TAC TTC TAATCATTTG TTAGAAATAC 1167
Ser Val Gly Val Gly Leu Arg Val Tyr Phe
355 360

← oligo #4 →

ATTATTAAAA GCAAGGCGAA TCGAAAGATT CGCTTTTTTT GCTCAAAATC AAGTTAAAAA 1227

AseI

ATGATTAAGT TAAAAGTGTA TAAATATTTA GGCTATTTTA TAAGTAACAA AATATTAATA 1287

← PCR-4 ← DraI

AAAAATCTGT GACATATATC ACAGATTTTT AAATCAATTA ACTATTTAAG TGTTTACTAT 1347

AseI ← PCR-5 ←

TAATTCTCTT TCCACTTTCC GTTTACTACT GTGCCGATTA CTTGGTAATT TGGCGTAAAC 1407

Sau3A

ACGGCTAAGT TTGCTATCTT ACCTTTTTCT ACCGAACCTA AACGATCATC TATACCAATT 1467

SalI

GCTCGTCGAC 1477

FIG.4C

```
    HdeI  NheI                                                    HindIII
CAT ATG GCT AGC ATG ACT GGT GGA CAG CAA ATG GGT CGG GAT TCA AGC        48
    Met Ala Ser Met Thr Gly Gly Gln Gln Met Gly Arg Asp Ser Ser
      1               5                  10                  15

    KpnI            BamHI
TTG GTA CCG AGC TCG GAT CCA GCT GTT GTT TAT AAC AAC GAA GGG ACT        96
Leu Val Pro Ser Ser Asp Pro Ala Val Val Tyr Asn Asn Glu Gly Thr
                20                  25                  30

AAC GTA GAA TTA GGT GGT CGT TTA AGC ATT ATC GCA GAA CAA AGT AAT       144
Asn Val Glu Leu Gly Gly Arg Leu Ser Ile Ile Ala Glu Gln Ser Asn
            35                  40                  45

AGC ACT GTA GAT AAT CAA AAA CAG CAA CAC GGT GCA TTA CGC AAT CAA       192
Ser Thr Val Asp Asn Gln Lys Gln Gln His Gly Ala Leu Arg Asn Gln
        50                  55                  60

GGT TCA CGT TTC CAC ATT AAA GCA ACT CAT AAC TTC GGT GAT GGT TTC       240
Gly Ser Arg Phe His Ile Lys Ala Thr His Asn Phe Gly Asp Gly Phe
    65                  70                  75

TAT GCA CAA GGT TAT TTA GAA ACT CGT TTT GTT ACA AAA GCC TCT GAA       288
Tyr Ala Gln Gly Tyr Leu Glu Thr Arg Phe Val Thr Lys Ala Ser Glu
80                  85                  90                  95

AAC GGT TCA GAT AAC TTC GGT GAT ATT ACA AGC AAA TAT GCT TAT GTT       336
Asn Gly Ser Asp Asn Phe Gly Asp Ile Thr Ser Lys Tyr Ala Tyr Val
                100                 105                 110

ACT TTA GGA AAT AAA GCA TTC GGT GAA GTA AAA CTT GGT CGT GCG AAA       384
Thr Leu Gly Asn Lys Ala Phe Gly Glu Val Lys Leu Gly Arg Ala Lys
            115                 120                 125

ACT ATT GCT GAT GGC ATA ACA AGT GCA GAA GAT AAA GAA TAT GGC GTT       432
Thr Ile Ala Asp Gly Ile Thr Ser Ala Glu Asp Lys Glu Tyr Gly Val
        130                 135                 140
```

FIG.5A

SpeI

CTC AAC AAT AGT GAC TAT ATT CCT ACT AGT GGT AAT ACG GTT GGC TAT 480
Leu Asn Asn Ser Asp Tyr Ile Pro Thr Ser Gly Asn Thr Val Gly Tyr
145 150 155

DraI HhaI

ACT TTT AAA GGT ATT GAT GGT TTA GTA TTA GGC GCT AAT TAT TTA TTA 528
Thr Phe Lys Gly Ile Asp Gly Leu Val Leu Gly Ala Asn Tyr Leu Leu
160 165 170 175

FnuDI

GCA CAA AAG CGT GAG GGT GCA AAA GGT GAA AAT AAG CGG CCT AAT GAT 576
Ala Gln Lys Arg Glu Gly Ala Lys Gly Glu Asn Lys Arg Pro Asn Asp
180 185 190

SnaBI EcoRI

AAG GCT GGT GAA GTA CGT ATA GGT GAA ATC AAT AAT GGA ATT CAA GTT 624
Lys Ala Gly Glu Val Arg Ile Gly Glu Ile Asn Asn Gly Ile Gln Val
195 200 205

GGT GCA AAA TAT GAT GCA AAC GAC ATC GTT GCA AAA ATT GCT TAT GGT 672
Gly Ala Lys Tyr Asp Ala Asn Asp Ile Val Ala Lys Ile Ala Tyr Gly
210 215 220

AGA ACT AAC TAC AAA TAT AAC GAA TCT GAC GAG CAT AAA CAG CAA TTA 720
Arg Thr Asn Tyr Lys Tyr Asn Glu Ser Asp Glu His Lys Gln Gln Leu
225 230 235

AAT GGT GTA TTA GCA ACT TTA GGC TAT CGT TTT AGT GAT TTA GGC TTA 768
Asn Gly Val Leu Ala Thr Leu Gly Tyr Arg Phe Ser Asp Leu Gly Leu
240 245 250 255

XbaI

TTA GTG TCT CTA GAT AGT GGC TAT GCA AAA ACT AAA AAC TAT AAA ATT 816
Leu Val Ser Leu Asp Ser Gly Tyr Ala Lys Thr Lys Asn Tyr Lys Ile
260 265 270

AseI

AAA CAC GAA AAA CGC TAT TTC GTA TCT CCA GGT TTC CAA TAT GAA TTA 864
Lys His Glu Lys Arg Tyr Phe Val Ser Pro Gly Phe Gln Tyr Glu Leu
275 280 285

FIG.5B

```
ATG GAA GAT ACT AAT GTC TAT GGC AAC TTC AAA TAT GAA CGC ACT TCT          912
Met Glu Asp Thr Asn Val Tyr Gly Asn Phe Lys Tyr Glu Arg Thr Ser
        290                 295                 300

GTA GAT CAA GGT GAA AAA ACA CGT GAA CAA GCA GTA TTA TTC GGT GTA          960
Val Asp Gln Gly Glu Lys Thr Arg Glu Gln Ala Val Leu Phe Gly Val
    305                 310                 315

GAT CAT AAA CTT CAC AAA CAA CTA TTA ACC TAT ATT GAA GGT GCT TAC         1008
Asp His Lys Leu His Lys Gln Leu Leu Thr Tyr Ile Glu Gly Ala Tyr
320                 325                 330                 335

GCT AGA ACT AGA ACA ACT GAG ACA GGT AAA GGC GTA AAA ACT GAA AAA         1056
Ala Arg Thr Arg Thr Thr Glu Thr Gly Lys Gly Val Lys Thr Glu Lys
                340                 345                 350

                        MluI
GAA AAA TCA GTG GGT GTA GGT TTA CGC GTT TAC TTC TAATCATTTG              1102
Glu Lys Ser Val Gly Val Gly Leu Arg Val Tyr Phe
            355                 360

                                XhoI
TTAGAAATAC ATTATTAAAA GCAAGGCGAC TCGAG .....                            1137
```

FIG.5C

```
    HdeI
CAT ATG GCT GTT GTT TAT AAC AAC GAA GGG ACT AAC GTA GAA TTA GGT      48
    Met Ala Val Val Tyr Asn Asn Glu Gly Thr Asn Val Glu Leu Gly
     1               5                  10                  15

GGT CGT TTA AGC ATT ATC GCA GAA CAA AGT AAT AGC ACT GTA GAT AAT      96
Gly Arg Leu Ser Ile Ile Ala Glu Gln Ser Asn Ser Thr Val Asp Asn
                20                  25                  30

CAA AAA CAG CAA CAC GGT GCA TTA CGC AAT CAA GGT TCA CGT TTC CAC     144
Gln Lys Gln Gln His Gly Ala Leu Arg Asn Gln Gly Ser Arg Phe His
            35                  40                  45

ATT AAA GCA ACT CAT AAC TTC GGT GAT GGT TTC TAT GCA CAA GGT TAT     192
Ile Lys Ala Thr His Asn Phe Gly Asp Gly Phe Tyr Ala Gln Gly Tyr
        50                  55                  60

TTA GAA ACT CGT TTT GTT ACA AAA GCC TCT GAA AAC GGT TCA GAT AAC     240
Leu Glu Thr Arg Phe Val Thr Lys Ala Ser Glu Asn Gly Ser Asp Asn
    65                  70                  75

TTC GGT GAT ATT ACA AGC AAA TAT GCT TAT GTT ACT TTA GGA AAT AAA     288
Phe Gly Asp Ile Thr Ser Lys Tyr Ala Tyr Val Thr Leu Gly Asn Lys
80                  85                  90                  95

GCA TTC GGT GAA GTA AAA CTT GGT CGT GCG AAA ACT ATT GCT GAT GGC     336
Ala Phe Gly Glu Val Lys Leu Gly Arg Ala Lys Thr Ile Ala Asp Gly
                100                 105                 110

ATA ACA AGT GCA GAA GAT AAA GAA TAT GGC GTT CTC AAC AAT AGT GAC     384
Ile Thr Ser Ala Glu Asp Lys Glu Tyr Gly Val Leu Asn Asn Ser Asp
            115                 120                 125

                SpeI                                DraI
TAT ATT CCT ACT AGT GGT AAT ACG GTT GGC TAT ACT TTT AAA GGT ATT     432
Tyr Ile Pro Thr Ser Gly Asn Thr Val Gly Tyr Thr Phe Lys Gly Ile
        130                 135                 140

                    HhaI
GAT GGT TTA GTA TTA GGC GCT AAT TAT TTA TTA GCA CAA AAG CGT GAG     480
Asp Gly Leu Val Leu Gly Ala Asn Tyr Leu Leu Ala Gln Lys Arg Glu
    145                 150                 155
```

FIG.6A

```
GGT GCA AAA GGT GAA AAT AAG CGG CCT AAT GAT AAG GCT GGT GAA GTA          528
Gly Ala Lys Gly Glu Asn Lys Arg Pro Asn Asp Lys Ala Gly Glu Val
160                 165                 170                 175
                                  EcoRI
CGT ATA GGT GAA ATC AAT AAT GGA ATT CAA GTT GGT GCA AAA TAT GAT          576
Arg Ile Gly Glu Ile Asn Asn Gly Ile Gln Val Gly Ala Lys Tyr Asp
                180                 185                 190

GCA AAC GAC ATC GTT GCA AAA ATT GCT TAT GGT AGA ACT AAC TAC AAA          624
Ala Asn Asp Ile Val Ala Lys Ile Ala Tyr Gly Arg Thr Asn Tyr Lys
            195                 200                 205

TAT AAC GAA TCT GAC GAG CAT AAA CAG CAA TTA AAT GGT GTA TTA GCA          672
Tyr Asn Glu Ser Asp Glu His Lys Gln Gln Leu Asn Gly Val Leu Ala
        210                 215                 220
                                                      XbaI
ACT TTA GGC TAT CGT TTT AGT GAT TTA GGC TTA TTA GTG TCT CTA GAT          720
Thr Leu Gly Tyr Arg Phe Ser Asp Leu Gly Leu Leu Val Ser Leu Asp
    225                 230                 235

AGT GGC TAT GCA AAA ACT AAA AAC TAT AAA ATT AAA CAC GAA AAA CGC          768
Ser Gly Tyr Ala Lys Thr Lys Asn Tyr Lys Ile Lys His Glu Lys Arg
240                 245                 250                 255
                                      AseI
TAT TTC GTA TCT CCA GGT TTC CAA TAT GAA TTA ATG GAA GAT ACT AAT          816
Tyr Phe Val Ser Pro Gly Phe Gln Tyr Glu Leu Met Glu Asp Thr Asn
                260                 265                 270

GTC TAT GGC AAC TTC AAA TAT GAA CGC ACT TCT GTA GAT CAA GGT GAA          864
Val Tyr Gly Asn Phe Lys Tyr Glu Arg Thr Ser Val Asp Gln Gly Glu
            275                 280                 285

AAA ACA CGT GAA CAA GCA GTA TTA TTC GGT GTA GAT CAT AAA CTT CAC          912
Lys Thr Arg Glu Gln Ala Val Leu Phe Gly Val Asp His Lys Leu His
        290                 295                 300

AAA CAA CTA TTA ACC TAT ATT GAA GGT GCT TAC GCT AGA ACT AGA ACA          960
Lys Gln Leu Leu Thr Tyr Ile Glu Gly Ala Tyr Ala Arg Thr Arg Thr
    305                 310                 315
```

FIG.6B

```
ACT GAG ACA GGT AAA GGC GTA AAA ACT GAA AAA GAA AAA TCA GTG GGT      1008

Thr Glu Thr Gly Lys Gly Val Lys Thr Glu Lys Glu Lys Ser Val Gly
320                 325                 330                 335

            MluI
GTA GGT TTA CGC GTT TAC TTC TAATCATTTG TTAGAAATAC ATTATTAAAA      1059
Val Gly Leu Arg Val Tyr Phe end
                340

            XhoI
GCAAGGCGAC TCGAG .....                                           1074
```

FIG.6C

```
    HdeI
CAT ATG GCT GTT GTT TAT AAC AAC GAA GGG ACT AAC GTA GAA TTA GGT      48
    Met Ala Val Val Tyr Asn Asn Glu Gly Thr Asn Val Glu Leu Gly
      1               5                  10                  15

GGT CGT TTA AGC ATT ATC GCA GAA CAA AGT AAT AGC ACT GTA GAT AAT      96
Gly Arg Leu Ser Ile Ile Ala Glu Gln Ser Asn Ser Thr Val Asp Asn
                  20                  25                  30

CAA AAA CAG CAA CAC GGT GCA TTA CGC AAT CAA GGT TCA CGT TTC CAC     144
Gln Lys Gln Gln His Gly Ala Leu Arg Asn Gln Gly Ser Arg Phe His
            35                  40                  45

ATT AAA GCA ACT CAT AAC TTC GGT GAT GGT TTC TAT GCA CAA GGT TAT     192
Ile Lys Ala Thr His Asn Phe Gly Asp Gly Phe Tyr Ala Gln Gly Tyr
        50                  55                  60

TTA GAA ACT CGT TTT GTT ACA AAA GCC TCT GAA AAC GGT TCA GAT AAC     240
Leu Glu Thr Arg Phe Val Thr Lys Ala Ser Glu Asn Gly Ser Asp Asn
    65                  70                  75

TTC GGT GAT ATT ACA AGC AAA TAT GCT TAT GTT ACT TTA GGA AAT AAA     288
Phe Gly Asp Ile Thr Ser Lys Tyr Ala Tyr Val Thr Leu Gly Asn Lys
80                  85                  90                  95

GCA TTC GGT GAA GTA AAA CTT GGT CGT GCG AAA ACT ATT GCT GAT GGC     336
Ala Phe Gly Glu Val Lys Leu Gly Arg Ala Lys Thr Ile Ala Asp Gly
                100                 105                 110

ATA ACA AGT GCA GAA GAT AAA GAA TAT GGC GTT CTC AAC AAT AGT GAC     384
Ile Thr Ser Ala Glu Asp Lys Glu Tyr Gly Val Leu Asn Asn Ser Asp
            115                 120                 125

              SpeI                                DraI
TAT ATT CCT ACT AGT GGT AAT ACG GTT GGC TAT ACT TTT AAA GGT ATT     432
Tyr Ile Pro Thr Ser Gly Asn Thr Val Gly Tyr Thr Phe Lys Gly Ile
        130                 135                 140
```

FIG.7A

HhaI

```
GAT GGT TTA GTA TTA GGC GCT AAT TAT TTA TTA GCA CAA AAG CGT GAG      480
Asp Gly Leu Val Leu Gly Ala Asn Tyr Leu Leu Ala Gln Lys Arg Glu
    145                 150                 155

GGT GCA AAA GGT GAA AAT AAG CGG CCT AAT GAT AAG GCT GGT GAA GTA      528
Gly Ala Lys Gly Glu Asn Lys Arg Pro Asn Asp Lys Ala Gly Glu Val
160                 165                 170                 175
```

EcoRI

```
CGT ATA GGT GAA ATC AAT AAT GGA ATT CAA GTT GGT GCA AAA TAT GAT      576
Arg Ile Gly Glu Ile Asn Asn Gly Ile Gln Val Gly Ala Lys Tyr Asp
                180                 185                 190

GCA AAC GAC ATC GTT GCA AAA ATT GCT TAT GGT AGA ACT AAC TAC AAA      624
Ala Asn Asp Ile Val Ala Lys Ile Ala Tyr Gly Arg Thr Asn Tyr Lys
            195                 200                 205

TAT AAC GAA TCT GAC GAG CAT AAA CAG CAA TTA AAT GGT GTA TTA GCA      672
Tyr Asn Glu Ser Asp Glu His Lys Gln Gln Leu Asn Gly Val Leu Ala
        210                 215                 220
```

XbaI

```
ACT TTA GGC TAT CGT TTT AGT GAT TTA GGC TTA TTA GTG TCT CTA GAT      720
Thr Leu Gly Tyr Arg Phe Ser Asp Leu Gly Leu Leu Val Ser Leu Asp
    225                 230                 235

AGT GGC TAT GCA AAA ACT AAA AAC TAT AAA ATT AAA CAC GAA AAA CGC      768
Ser Gly Tyr Ala Lys Thr Lys Asn Tyr Lys Ile Lys His Glu Lys Arg
240                 245                 250                 255
```

AseI

```
TAT TTC GTA TCT CCA GGT TTC CAA TAT GAA TTA ATG GAA GAT ACT AAT      816
Tyr Phe Val Ser Pro Gly Phe Gln Tyr Glu Leu Met Glu Asp Thr Asn
                260                 265                 270

GTC TAT GGC AAC TTC AAA TAT GAA CGC ACT TCT GTA GAT CAA GGT GAA      864
Val Tyr Gly Asn Phe Lys Tyr Glu Arg Thr Ser Val Asp Gln Gly Glu
            275                 280                 285

AAA ACA CGT GAA CAA GCA GTA TTA TTC GGT GTA GAT CAT AAA CTT CAC      912
Lys Thr Arg Glu Gln Ala Val Leu Phe Gly Val Asp His Lys Leu His
```

FIG.7B

```
        290                 295                 300

AAA CAA CTA TTA ACC TAT ATT GAA GGT GCT TAC GCT AGA ACT AGA ACA      960
Lys Gln Leu Leu Thr Tyr Ile Glu Gly Ala Tyr Ala Arg Thr Arg Thr
    305                 310                 315

ACT GAG ACA GGT AAA GGC GTA AAA ACT GAA AAA GAA AAA TCA GTG GGT     1008
Thr Glu Thr Gly Lys Gly Val Lys Thr Glu Lys Glu Lys Ser Val Gly
320                 325                 330                 335

            MluI
GTA GGT TTA CGC GTT TAC TTC TAATCATTTG TTAGAAATAC ATTATTAAAA        1059
Val Gly Leu Arg Val Tyr Phe
                340
        BamHI
GCAAGGCGGA TCC .....                                               1072
```

FIG.7C

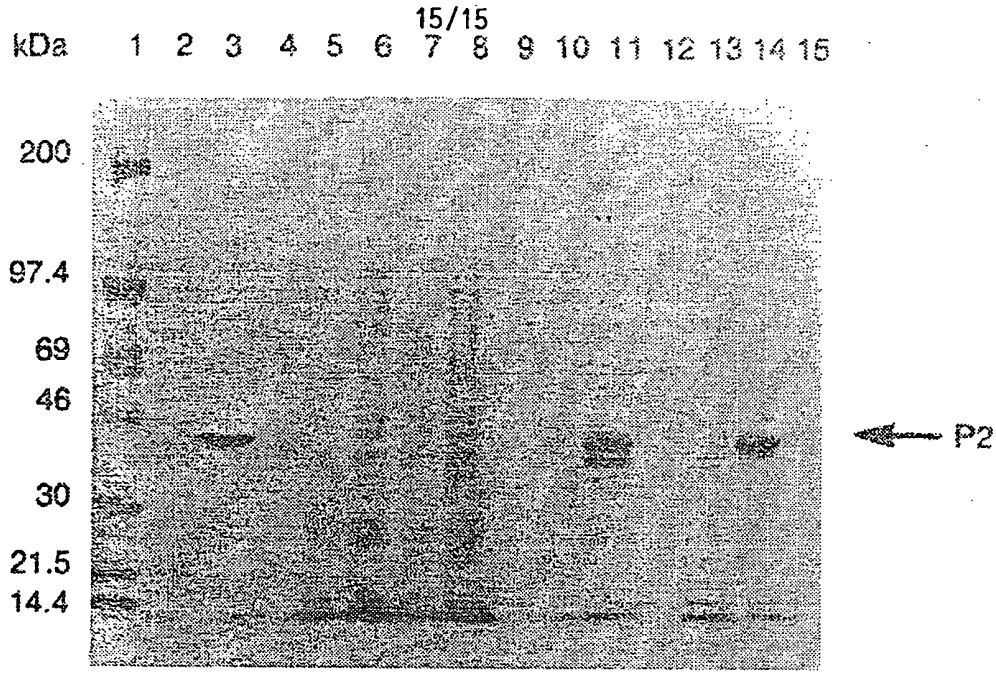

kDa 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
200
97.4
69
46
30
21.5
14.4
P2

FIG. 8A

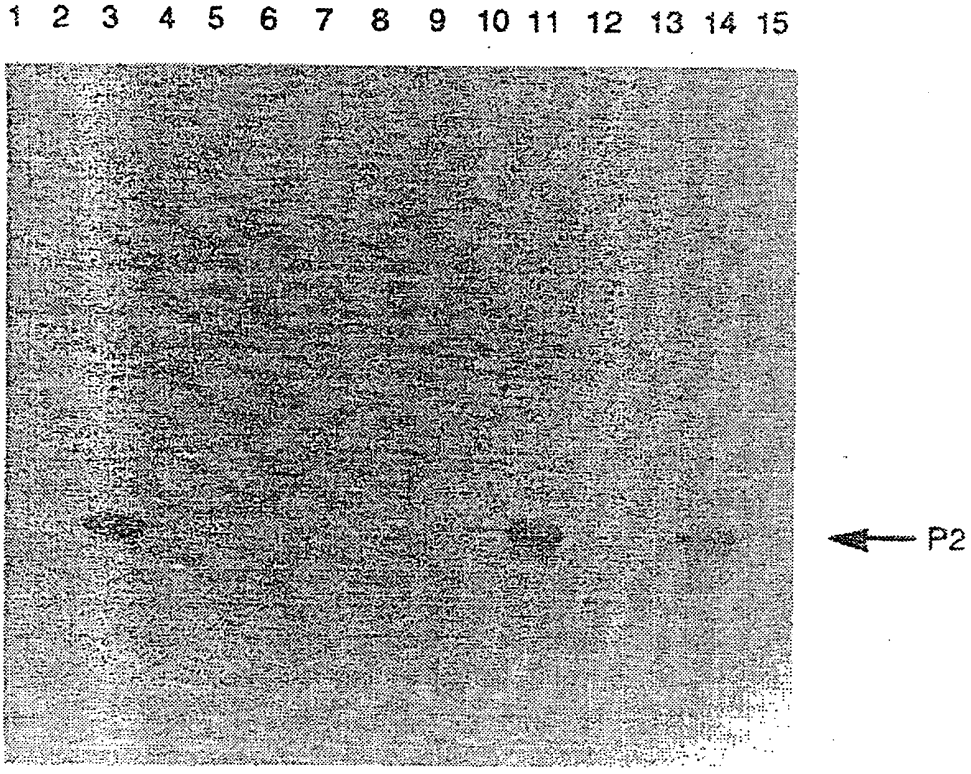

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
P2

FIG. 8B

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US94/08326

A. CLASSIFICATION OF SUBJECT MATTER

IPC(5) :Please See Extra Sheet.
US CL :Please See Extra Sheet.
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : Please See Extra Sheet.

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	EP, A, 0,320,289 (HANSEN) 14 JUNE 1989, see entire document.	6-11
Y	EP, A, 0,378,929 (MUNSON JR. ET AL.) 25 JULY 1990, see entire document.	6-11
Y	INFECTION AND IMMUNITY, Volume 57, Number 1, issued January 1989, R. Munson Jr. et al., "Molecular Cloning, Expression, and Primary Sequence of Outer Membrane P2 of Haemophilus Influenzae Type b", pages 88-94, see entire document.	1-3, 8-10, 12

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
19 SEPTEMBER 1994	27 OCT 1994
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer ANTHONY C. CAPUTA Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US94/08326

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X -- Y	JOURNAL OF CLINICAL INVESTIGATION, Volume 72, issued August 1983, R.S. Munson Jr. et al., "Purification and Comparison of Outer Membrane Protein P2 from Haemophilus influenzae Type b isolates", pages 677-684, see entire document.	8 --------- 6, 7, 9, 10-12
Y	JOURNAL OF EXPERIMENTAL MEDICINE, Volume 152, issued 1980, R. Schneerson et al., "Preparation, Characterization, and Immunogenicity of Haemophilus influenzae Type b Polysaccharide-Protein Conjugates", pages 361-376, see entire document.	6-11
Y	METHODS IN ENZYMOLOGY, Volume 182, issued 1990, F.O. Martson et al., "Solubilization of Protein Aggregates", pages 264-276, see entire document.	6-11
Y	US, A, 4,656,255 (SEELY) 07 April 1987, see entire document.	6-11
Y	JOURNAL OF BACTERIOLOGY, Volume 162, Number 3, issued June 1985, V. Vachon et al., "Transmembrane Permeability Channels Across the Outer Membrane of Haemophilus influenzae Type b", pages 918-924, see entire document.	6-11
Y	GIBCO BRL CATALOGUE AND REFERENCE GUIDE 1992, published 1991, pages 355 and 357, see pages 355 and 357.	1-5, 12-15
Y	METHODS IN ENZYMOLOGY, Volume 185, issued 1990, F.W. Studier et al., "Uses of T7 RNA Polymerase to Direct Expression of Cloned Genes", pages 60-89, see entire document.	1-5, 12-15
Y	INFECTION AND IMMUNITY, Volume 57, Number 4, issued April 1989, E.J. Hansen et al., "Primary Structure of the Porin Protein of Haemophilus Influenzae Type b Determined by Nucleotide Sequence Analysis", pages 1100-1107, see entire document.	1-15

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US94/08326

A. CLASSIFICATION OF SUBJECT MATTER:
IPC (5):

A61K 39/102; C07K 3/12, 3/18, 3/20, 3/26, 3/28, 15/04, 17/10; C12N 15/31, 15/70; C12P 21/02

A. CLASSIFICATION OF SUBJECT MATTER:
US CL :

424/256.1; 435/69.3, 320.1, 851; 530/350, 402, 412, 414, 415, 417

B. FIELDS SEARCHED
Minimum documentation searched
Classification System: U.S.

424/256.1; 435/69.3, 320.1, 851; 530/350, 402, 412, 414, 415, 417

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

CA, CABA, CAPREVIEWS, BIOSIS, MEDLINE, BIOTECHABS, BIOTECHDS, JICST-E, LIFESCI, WPIDS, IFIPAT, INPADOC, WPINDEX, DISSABS, GENBANK, CJACS, CJELSEVIER, PATOSDE, PATOSWO, PATOSEP, ANABSTR, AQUASCI, CEABA, CEN, CIN, FSTA, CONFSCI, DRUGNL. BIOBUSINESS
search terms: P2, TAI?, PULLEN?, SPOER?, LAING?, PET 16B, T7, CAPSID, PROTEIN. NOVAGEN?, PET 17B, PET?